161628

STIC EIC 2100
Search Request Form (117)

Today's Date: 8-1-5

What date would you like to use to limit the search?

Priority Date: 11/23/2000   Other:

Name HANA THAI

AU 2161   Examiner # 79864

Room # 3C29   Phone 2-4029

Serial # 09/989,970

Format for Search Results (Circle One):

(PAPER)   DISK   EMAIL

Where have you searched so far?

USP   DWPI   EPO   JPO   ACM   IBM TDB

IEEE   INSPEC   SPI   Other _____

Is this a "Fast & Focused" Search Request? (Circle One) (YES)   NO

A "Fast & Focused" Search is completed in 2-3 hours (maximum). The search must be on a very specific topic and meet certain criteria. The criteria are posted in EIC2100 and on the EIC2100 NPL Web Page at http://ptoweb/patents/stic/stic-tc2100.htm.

What is the topic, novelty, motivation, utility, or other specific details defining the desired focus of this search? Please include the concepts, synonyms, keywords, acronyms, definitions, strategies, and anything else that helps to describe the topic. Please attach a copy of the abstract, background, brief summary, pertinent claims and any citations of relevant art you have found.

Determing whether relevant information is contained in the document

generating a document extract if the document contains relevant

information, extracting a portion of the document and

original

Replacing the document in the step with the document extract

Please see the attached claims.

STIC Searcher C-Wong   Phone 272-3513

Date picked up 8-7   Date Completed 8-1-05

sira

Search and Information
Resources Administration

File 696:DIALOG Telecom. Newsletters 1995-2005/Jul 29
        (c) 2005 Dialog
File 15:ABI/Inform(R) 1971-2005/Aug 01
        (c) 2005 ProQuest Info&Learning
File 98:General Sci Abs/Full-Text 1984-2004/Dec
        (c) 2005 The HW Wilson Co.
File 112:UBM Industry News 1998-2004/Jan 27
        (c) 2004 United Business Media
File 141:Readers Guide 1983-2004/Dec
        (c) 2005 The HW Wilson Co
File 484:Periodical Abs Plustext 1986-2005/Jul W4
        (c) 2005 ProQuest
File 608:KR/T Bus.News. 1992-2005/Aug 01
        (c)2005 Knight Ridder/Tribune Bus News
File 813:PR Newswire 1987-1999/Apr 30
        (c) 1999 PR Newswire Association Inc
File 613:PR Newswire 1999-2005/Aug 01
        (c) 2005 PR Newswire Association Inc
File 635:Business Dateline(R) 1985-2005/Jul 30
        (c) 2005 ProQuest Info&Learning
File 810:Business Wire 1986-1999/Feb 28
        (c) 1999 Business Wire
File 610:Business Wire 1999-2005/Aug 01
        (c) 2005 Business Wire.
File 369:New Scientist 1994-2005/May W4
        (c) 2005 Reed Business Information Ltd.
File 370:Science 1996-1999/Jul W3
        (c) 1999 AAAS
File 20:Dialog Global Reporter 1997-2005/Aug 01
        (c) 2005 Dialog
File 624:McGraw-Hill Publications 1985-2005/Aug 01
        (c) 2005 McGraw-Hill Co. Inc
File 634:San Jose Mercury Jun 1985-2005/Jul 30
        (c) 2005 San Jose Mercury News
File 647:CMP Computer Fulltext 1988-2005/Jul W3
        (c) 2005 CMP Media, LLC
File 674:Computer News Fulltext 1989-2005/Jul W4
        (c) 2005 IDG Communications

Set   Items   Description
S1   225558 DISPLAC? OR SUBROGAT?
S2   9831039 REPLAC??? ? OR REPLACEMENT? OR SUBSTITUT???? ? OR RE()(PLA-
                C??? ? OR PLACEMENT? ?) OR EXCHANG? OR SWAP? OR SWITCH? OR SU-
                PPLANT?
S3   107461 FULLTEXT? OR (FULL OR COMPLETE)()TEXT
S4   16162575 DOCUMENT? ? OR ARTICLE? ? OR RECORD? ? OR REPORT? ? OR MES-
                SAGE? ?
S5   6537189 ABSTRACT? ? OR SUMMARY? OR SUMMARIES OR SUMMATION? OR SHOR-
                TEN? OR SHORTER? OR ABBREVIAT? OR REDUC??? ? OR REDUCTION?
S6   2340858 ABRIDG? OR CONDENS??? ? OR PRECIS OR SYNOPSI? OR CAPSUL? OR
                RECAP? ? OR BRIEF?? ? OR DIGEST? ?
S7   192679 EXTRACT? ?
S8   685688 S1:S2(10N)S3:S4
S9   10935   S8(10N)S5:S7
S10   6275314 SEARCH? OR RETRIEV? OR HARVEST? OR QUERY? OR QUERIE? OR MI-
                NE? ? OR MINING OR DATAMIN? OR TEXTSEARCH? OR REQUEST?
S11   363430 IR
S12   1131   S11(3N)S4
S13   2   S9(S)S12
S14   357   S9(S)S10:S11

```
S15    434498   DATA()(BASE? ? OR SET? ? OR BANK? ? OR STORE? OR SYSTEM? ?
                 OR COLLECTION? OR LIBRAR? OR DEPOSIT? OR REPOSIT? OR WAREHOUS-
                 E? OR WARE? ? OR MART? ?)
S16      3057   DATAWAREHOUS? OR DATAMART?
S17   1092373   DATABASE? OR DATASET? OR DATABANK? OR DATASTORE? OR DATASY-
                 STEM? OR DATACOLLECTION? OR DATALIBRAR? OR DATADEPOSIT? OR DA-
                 TAREPOSIT?
S18       167   S9(S)S15:S17
S19    345734   S1:S2(5N)S3:S4
S20      2789   S19(5N)S5:S7
S21       120   S20(S)S10:S11
S22        58   S20(S)S15:S17
S23        28   S21 AND S22
S24         5   S23/2001:2005
S25        23   S23 NOT S24
S26        19   RD (unique items)
S27   3312880   REPLAC??? ? OR REPLACEMENT? OR SUBSTITUT? OR RE()(PLAC??? ?
                 OR PLACEMENT? ?) OR SWAP? OR SUPPLANT?
S28     34124   (S1 OR S27)(5N)S3:S4
S29      2148   S28(S)S5:S7
S30       236   S29(S)S10:S11
S31       100   S29(S)S15:S17
S32       102   S31 OR S13
S33        41   S32/2001:2005
S34        52   S32 NOT (S33 OR S23)
S35        46   RD (unique items)
```

35/3,K/3   (Item 3 from file: 15)
DIALOG(R)File   15:ABI/Inform(R)

02070592 61051647
ETL provides the keys to keeping data relevant, available
Steinacher, Scott
InfoWorld v22n39 PP: 74-76 Sep 25, 2000
ISSN: 0199-6649 JRNL CODE: IFW
WORD COUNT: 1305

...TEXT: is preferable because it reduces response times and enhances ease
of use.

Before loading a **data   mart** , programmers typically aggregate data.
Aggregation routines **replace** numerous detail **records** with relatively
few **summary** records. For example, suppose that a year's worth of sales
data is stored in several thousand records in a normalized **database** .

Through aggregation, this data is transformed into fewer summary records
that will be written to...

35/3,K/6   (Item 6 from file: 15)
DIALOG(R)File   15:ABI/Inform(R)

01539058 01-90046
PubMed initiates new MEDLINE era
O Leary, Mick
Database v20n6 PP: 70-72 Dec 1997
ISSN: 0162-4105 JRNL CODE: DTB
WORD COUNT: 1436

...TEXT: other Web medical data. And it is free.

MEDLINE PLUS

PubMed has the complete MEDLINE **database** , from 1966 forward. It is in a single file (though date limiting is available), and is updated weekly. It includes PreMEDLINE records, which contain citation and **abstract** , but which have not been indexed. These are added daily, and are eventually **replaced** by fully indexed **records** .

PubMed also has a small class of records which do not appear in other versions...

File 348:EUROPEAN PATENTS 1978-2005/Jul W04
        (c) 2005 European Patent Office
File 349:PCT FULLTEXT 1979-2005/UB=20050728,UT=20050721
        (c) 2005 WIPO/Univentio
File 324:German Patents Fulltext 1967-200529
        (c) 2005 Univentio

Set	Items	Description
S1	349527	DISPLAC? OR SUBROGAT?
S2	1909686	REPLAC??? ? OR REPLACEMENT? OR SUBSTITUT???? ? OR RE()(PLAC??? ? OR PLACEMENT? ?) OR EXCHANG? OR SWAP? OR SWITCH? OR SUPPLANT?
S3	1704	FULLTEXT? OR (FULL OR COMPLETE)()TEXT
S4	2520974	DOCUMENT? ? OR ARTICLE? ? OR RECORD? ? OR REPORT? ? OR MESSAGE? ?
S5	2153694	ABSTRACT? ? OR SUMMARY? OR SUMMARIES OR SUMMATION? OR SHORTEN? OR SHORTER? OR ABBREVIAT? OR REDUC??? ? OR REDUCTION?
S6	1235167	ABRIDG? OR CONDENS??? ? OR PRECIS OR SYNOPSI? OR CAPSUL? OR RECAP? ? OR BRIEF?? ? OR DIGEST? ?
S7	211399	EXTRACT? ?
S8	85253	S1:S2(10N)S3:S4
S9	2798	S8(10N)S5:S7
S10	2363105	SEARCH? OR RETRIEV? OR HARVEST? OR QUERY? OR QUERIE? OR MINE? ? OR MINING OR DATAMIN? OR TEXTSEARCH? OR REQUEST?
S11	189274	IR
S12	2541	S11(3N)S4
S13	7	S9(20N)S11:S12
S14	2633	S8(10N)S5:S6
S15	126	S9/TI,AB
S16	1689	IC='G06F-007/00'
S17	6	S9 AND S16
S18	42	S15 AND AC=US/PR
S19	36	S18 AND AY=(1976:2000)/PR
S20	102	S15 AND PY=1976:2000
S21	36	S19 NOT (S13 OR S17)
S22	36	IDPAT (sorted in duplicate/non-duplicate order)
S23	36	IDPAT (primary/non-duplicate records only)
S24	56	S20 NOT SWITCH?
S25	38	S24 NOT (S13 OR S17 OR S23)
S26	38	IDPAT (sorted in duplicate/non-duplicate order)
S27	37	IDPAT (primary/non-duplicate records only)
S28	184466	IC=G06F
S29	424	S14 AND S28
S30	48	S29/TI,AB,CM
S31	38	S30 NOT (S13 OR S17 OR S23 OR S15)
S32	109088	DATABASE? OR DATASET? OR DATABANK? OR DATASTORE? OR DATASYSTEM? OR DATACOLLECTION? OR DATALIBRAR? OR DATADEPOSIT? OR DATAREPOSIT?
S33	79	DATAWAREHOUSE? OR DATAMART?
S34	195617	DATA()(BASE? ? OR SET? ? OR BANK? ? OR STORE? OR SYSTEM? ? OR COLLECTION? OR LIBRAR? OR DEPOSIT? OR REPOSIT? OR WAREHOUSE? OR WARE? ? OR MART? ?)
S35	17402	IC='G06F-017/30':IC='G06F-017/32'
S36	77	S14 AND S35
S37	194	S9(20N)S10:S11
S38	58	S37 NOT SWITCH?
S39	66	S14(20N)S32:S34
S40	171	S36 OR S38:S39
S41	141	S40 NOT (S13 OR S17 OR S23 OR S15 OR S25 OR S31)
S42	112	S41 AND AC=US/PR
S43	87	S42 AND AY=(1976:2000)/PR

```
S44        67    S41 AND PY=1976:2000
S45        99    S43:S44
S46        99    IDPAT (sorted in duplicate/non-duplicate order)
S47        99    IDPAT (primary/non-duplicate records only)
?
```

00574703   **Image available**
ABBREVIATING AND CONDENSING TEXT IN COMPUTER SOFTWARE
ABREVIATION ET CONDENSATION DE TEXTE DANS UN LOGICIEL
Patent Applicant/Assignee:
  KUDROLLIS SOFTWARE INVENTIONS PVT LTD,
Inventor(s):
  KUDROLLI Abdus Samad,
  KUDROLLI Parvez,
  KUDROLLI Feroz,
Patent and Priority Information (Country, Number, Date):
  Patent: WO 200038076 A1 20000629 (WO 0038076)
  Application: WO 99IN64 19991116 (PCT/WO IN9900064)
  Priority Application: IN 98827 19981221
Designated States:
(Protection type is "patent" unless otherwise stated - for applications
prior to 2004)
  CA AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE
Main International Patent Class: G06F-015/00
International Patent Class: **G06F-007/00** ; G06F-017/00; G06F-017/21;
  G06F-017/24; G06F-017/27
Publication Language: English
Fulltext Availability:
  Detailed Description
  Claims
Fulltext Word Count: 28220

English Abstract
   This invention relates to text abbreviation methods in computer
   software. In particular, abbreviation of text into predetermined field
   widths (with single or multiple rows) or files, utilizing an operating
   system (121), an application program (122), and an abbreviation control
   data program (123), along with combinations of prioritized shortening
   methods in preference to or in addition to glossaries of acronyms and
   word abbreviations using an abbreviation function (127) are disclosed.
   The special handling of segments of input contained within pairs of
   pre-defined characters, as well as omission of spaces, and conversion of
   enumeration word or word sequences to numbers utilizing an abbreviation
   data file (124), a parameters sets file (125), and a parameters list
   (126), are also disclosed. The omission of spaces and phonetically less
   significant characters condenses word sequences, which saves display
   space and enables the use of larger type sizes.

French Abstract
   L'invention concerne des procedes d'abreviation de texte dans un
   logiciel. L'invention concerne plus particulierement l'abreviation de
   texte selon des largeurs de zones predeterminees (avec des rangees
   simples ou multiples) ou de fichiers, a l'aide d'un systeme
   d'exploitation (121), d'un programme d'application (122) et d'un
   programme de donnees de commande d'abreviation (123) ainsi que des
   combinaisons de procedes permetttant de raccourcir le texte preferes au
   detriment de ou en sus de glossaires classiques d'acronymes et
   d'abreviation de mots grace a une fonction d'abreviation (127). Ces
   procedes consistent a manipuler des segments d'entree se trouvant dans
   des paires de caracteres predefinis, a omettre les espaces ou la
   ponctuation predefinie, a convertir une enumeration de mots ou une
   sequence de mots en nombres utilisant un fichier donnees d'abreviation

(124). L'invention concerne aussi un fichier de jeux de parametres (125) et une liste de parametres (126). L'omission d'espaces et des caracteres moins signifiants en termes de phonetique permettent d'economiser de l'espace d'affichage et d'utiliser des equipements de dimensions plus grandes.

International Patent Class: **G06F-007/00** ...
Fulltext Availability:
  Detailed Description

Detailed Description
...   file 1282 to keep
  5 track of reduction scope length of the acronyms or word
   abbreviations found and held for need based **replacement** . The
   records in **reduction**   scope file are sequenced in the
  descending order of reduction scope length, the objective
  being to achieve the required **reduction** with the least number
  10 of need based **replacements** in the **records** of the   **shortening**
  file as referenced from the first few records of the
  reduction scope file.

  If the...common Match subroutine
  (Method 29) is called for compulsory and need based acronym
  and word **abbreviation**   search, **replacement** or retention 15a.

  In file:Shrtn, if a **record** is found with field:ShSWrd=,And,
  and field:ShSWrd in preceding and succeeding records does...

  ...the following specific steps,
  designated (a) to (z), together with relevant explanations
  and notes.

  a) **Replacing** enumeration words with ' **abbreviations** .

  Accessing file:Shrtn and reading each **record** .

  If ShRS=1 accessing file:AbData and locating and reading
  any record with AbAR=2...replaced in upper case
  to avoid confusion with subsequent abbreviation methods.

  Before proceeding further in **Shorten** subroutine it is
  necessary to search for **records** having sequences **replaced** in
  upper case with ShCap=10 or 11 and to convert them back to
  - 76...
?

DIALOG(R)File 348:EUROPEAN PATENTS
(c) 2005 European Patent Office. All rts. reserv.

00957350
**Method for transmitting alternative messages in a TDMA system with
    discontinuous transmission**
**Verfahren zur Übertragung alternativer Nachrichten in einem TDMA-System mit
    diskontinuierlicher Übertragung**
**Methode de transmission de messages alternatifs dans un systeme AMRT a
    transmission discontinue**
PATENT ASSIGNEE:
  LUCENT TECHNOLOGIES INC., (2143720), 600 Mountain Avenue, Murray Hill,
    New Jersey 07974-0636, (US), (Proprietor designated states: all)
INVENTOR:
  Kahn, Colin Leon, 38 Cross Road, Cedar Knolls, New Jersey 07950, (US)
  Turner, Michael D., 33 Elm No. 9, Madison, New Jersey 07940, (US)
LEGAL REPRESENTATIVE:
  Buckley, Christopher Simon Thirsk et al (28912), Lucent Technologies (UK)
    Ltd, 5 Mornington Road, Woodford Green, Essex IG8 0TU, (GB)
PATENT (CC, No, Kind, Date): EP 868037 A1 980930 (Basic)
                              EP 868037 B1 020320
APPLICATION (CC, No, Date):   EP 98301801 980311;
PRIORITY (CC, No, Date): US 821479 970321
DESIGNATED STATES: DE; FI; FR; GB; SE
INTERNATIONAL PATENT CLASS: H04B-007/26
CITED PATENTS (EP B): WO 91/02436 A; WO 96/33585 A; GB 2290198 A; US
    5581548 A

ABSTRACT EP 868037 A1
    Alternative messages are transmitted in a time slot (e.g.3) of a frame
  of a time division multiple access communications system during periods
  of silence or when no speech data is present. When an absence of voice is
  detected, an **abbreviated   message** (60) is **substituted** for a longer
  message in the time slot (31). In subsequent frames, the **shorter
  message** (60) is also **substituted** for the longer **message** until voice
  is detected.
ABSTRACT WORD COUNT: 72
NOTE:
  Figure number on first page: 3

LEGAL STATUS (Type, Pub Date, Kind, Text):
 Change: 000913 A1 Title of invention (German) changed: 20000724
 Application: 980930 A1 Published application (A1with Search Report
                           ;A2without Search Report)
 Lapse: 030723 B1 Date of lapse of European Patent in a
                           contracting state (Country, date): DE
                           20020621, FI 20020320, SE 20020620,
 Lapse: 021204 B1 Date of lapse of European Patent in a
                           contracting state (Country, date): FI
                           20020320, SE 20020620,
 Grant: 020320 B1 Granted patent
 Change: 000913 A1 Title of invention (English) changed: 20000724
 Change: 000913 A1 Title of invention (French) changed: 20000724
 Lapse: 020911 B1 Date of lapse of European Patent in a
                           contracting state (Country, date): SE
                           20020620,
 Oppn None: 030312 B1 No opposition filed: 20021223

Examination:	980930 A1 Date of filing of request for examination: 980320
Examination:	981125 A1 Date of despatch of first examination report: 981008
Change:	990414 A1 Title of invention (German) (change)
Change:	990414 A1 Title of invention (English) (change)
Change:	990414 A1 Title of invention (French) (change)
Change:	990609 A1 Designated Contracting States (change)

LANGUAGE (Publication,Procedural,Application): English; English; English
FULLTEXT AVAILABILITY:

Available Text	Language	Update	Word Count
CLAIMS A	(English)	199840	418
CLAIMS B	(English)	200212	235
CLAIMS B	(German)	200212	210
CLAIMS B	(French)	200212	294
SPEC A	(English)	199840	1754
SPEC B	(English)	200212	1934
Total word count - document A			2172
Total word count - document B			2673
Total word count - documents A + B			4845

...ABSTRACT or when no speech data is present. When an absence of voice is
detected, an **abbreviated message** (60) is **substituted** for a longer
message in the time slot (31). In subsequent frames, the **shorter
message** (60) is also **substituted** for the longer **message** until voice
is detected.

23/5,K/25   (Item 25 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT
(c) 2005 WIPO/Univentio. All rts. reserv.

00878841   **Image available**
**METHOD AND SYSTEM FOR ACCELERATING THE DELIVERY OF CONTENT IN A NETWORKED
    ENVIRONMENT**
**PROCEDE ET SYSTEME POUR ACCELERER L'EXPEDITION DES CONTENUS DANS UN
    ENVIRONNEMENT RESEAU**
Patent Applicant/Assignee:
  FINEGROUND NETWORKS, Suite 201, 42 West Campbell Avenue, Campbell, CA
    95008-1042, US, US (Residence), US (Nationality)
Inventor(s):
  KAUSIK Balas Natarajan, 18079 Reed Knoll Road, Los Gatos, CA 95033, US,
Legal Representative:
  YANG Joseph (et al) (agent), Skadden, Arps, Slate, Meagher & Flom LLP,
    525 University Avenue, Palo Alto, CA 94301, US,
Patent and Priority Information (Country, Number, Date):
  Patent: WO 200213037 A1 20020214 (WO 0213037)
  Application: WO 2001US24936 20010808 (PCT/WO US0124936)
  Priority Application: US 2000634134 20000808
Designated States:
(Protection type is "patent" unless otherwise stated - for applications
prior to 2004)
  AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ
  EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR
  LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL
  TJ TM TR TT TZ UA UG UZ VN YU ZA ZW
  (EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR
  (OA) BF BJ CF CG CI CM GA GN GQ GW ML MR NE SN TD TG
  (AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZW
  (EA) AM AZ BY KG KZ MD RU TJ TM
Main International Patent Class: G06F-015/16

English Abstract
  Many documents transmitted in a network environment contain substantial
  overlap with old versions of the same (or related) documents. For
  example, "current news" web page may be updated hourly on a web site so
  that a new story is added and the oldest story is dropped. In such cases,
  it in inefficient to send the updated document in its entirety to a user
  requesting the new document but who had previously received the old page.
  Instead, the new **document** is first sent to a **condenser** (200), which
  replaces the unchanged portions of the new **document** with pointers to
  the old document. In this way, only the changed portions of the document
  need to be transmitted to the user in their entirety. The condensed
  document is bound to the requesting user via a token such as a cookie
  generated by the condenser, and the condensed document and cookie are
  sent to the user. The user uses assembly software corresponding to the
  condensing process to reassemble the new document from the condensed
  document and the old document. The foregoing may be implemented on an
  individual user basis, as well as for classes of users.

French Abstract
  Beaucoup de documents transmis dans un environnement reseau et les
  anciennes versions de ces memes documents (ou de documents en rapport
  avec ceux-ci) se recoupent sensiblement. Par exemple, la page Web d'<=
  informations d'actualite >= d'un site Web peut etre mise a jour toutes
  les heures, afin d'ajouter une nouvelle histoire qui remplace la plus
  ancienne. Dans ce cas, il est inefficace d'expedier la totalite du
  document mis a jour a un utilisateur l'ayant sollicite, s'il a deja recu
  auparavant l'ancienne page. Au lieu de cela, le nouveau document Web est
  d'abord envoye a un condenseur (200), qui remplace les parties inchangees
  du nouveau document par des pointeurs renvoyant a l'ancien document.
  Ainsi, seules les parties modifiees du document doivent etre entierement
  transmises a l'utilisateur. Le document condense est mis en rapport avec
  l'utilisateur l'ayant sollicite au moyen d'un jeton d'identification, tel
  qu'un petit gateau genere par le condenseur, qui lui est expedie avec le
  document condense. L'utilisateur emploie un programme d'assemblage
  correspondant au procede permettant de condenser pour rassembler le
  nouveau document a partir du document condense et de l'ancien document.
  Le procede decrit peut etre applique pour un utilisateur individuel ou
  pour des categories d'utilisateurs.

English Abstract
  ...requesting the new document but who had previously received the old
  page. Instead, the new **document** is first sent to a **condenser** (200),
  which **replaces** the unchanged portions of the new **document** with
  pointers to the old document. In this way, only the changed portions of
  the...
?

? t47/5,k/40,42,54

47/5,K/40   (Item 40 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT
(c) 2005 WIPO/Univentio. All rts. reserv.

00910813   **Image available**
INTERACTIVE DISPLAY OF A DOCUMENT SUMMARY
AFFICHAGE INTERACTIF D'UN RESUME DE DOCUMENT
Patent Applicant/Assignee:
  FIRESPOUT INC, 448 Common Street, Belmont, MA 02478, US, US (Residence),
    US (Nationality), (For all designated states except: US)
Patent Applicant/Inventor:
  ARMSTRONG Juliet, 70 Park Street #64, Somerville, MA 02143, US, US
    (Residence), US (Nationality), (Designated only for: US)
  WEST David, 18 Union Street, #1, Cambridge, MA 02141, US, US (Residence),
    US (Nationality), (Designated only for: US)
  MILBERT Randy L, 339 Broadway Street, #3, Cambridge, MA 02139, US, US
    (Residence), US (Nationality), (Designated only for: US)
  VU Sonny, 641 Green Street, Cambridge, MA 02139, US, US (Residence), US
    (Nationality), (Designated only for: US)
  DAO Ngon D, 76 Fifth Street, #2, Cambridge, MA 02141, US, US (Residence),
    (Designated only for: US)
Legal Representative:
  LICHAUCO Faustino A (agent), Fish & Richardson, P.C., 225 Franklin
    Street, Boston, MA 02110-2804, US,
Patent and Priority Information (Country, Number, Date):
  Patent: WO 200244948 A2-A3 20020606 (WO 0244948)
  Application: WO 2001US44375 20011128   (PCT/WO US01044375)
  Priority Application: US 2000723966 20001128
Parent Application/Grant:
  Related by Continuation to: US 2000723966 20001128 (CON)
Designated States:
(Protection type is "patent" unless otherwise stated - for applications
prior to 2004)
  AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ
  EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR
  LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL
  TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
  (EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR
  (OA) BF BJ CF CG CI CM GA GN GQ GW ML MR NE SN TD TG
  (AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZM ZW
  (EA) AM AZ BY KG KZ MD RU TJ TM
Main International Patent Class: **G06F-017/30**
Publication Language: English
Filing Language: English
Fulltext Availability:
  Detailed Description
  Claims
Fulltext Word Count: 6590

English Abstract
  A method for displaying information indicative of the content of a target
  document of a link displayed in a first window includes detecting an
  event indicative of a user's interest in the target document. The user's
  interest can be indicated by moving a mouse pointer into an active region
  associated with the link. Upon detecting this event, information
  indicative of the content of the target document is retrieved and
  presented to the user without having to open the target document. One
  mechanism for presenting the information is to open a second window and

to display the information in the second window.

French Abstract
  Procede d'affichage d'informations revelatrices du contenu d'un document
  cible d'un lien affiche dans une premiere fenetre. Ledit procede comporte
  la detection d'un evenement indicateur d'un interet de l'utilisateur dans
  le document cible. L'interet de l'utilisateur peut etre indique par un
  deplacement du pointeur de la souris dans une region active associee au
  lien. Une fois cet evenement detecte, des informations revelatrices du
  contenu du document cible sont extraites et presentees a l'utilisateur
  sans que ce dernier ait a ouvrir le document cible. Dans un mecanisme de
  presentation des informations, une seconde fenetre est ouverte et les
  informations sont affichees dans ladite seconde fenetre.

Legal Status (Type, Date, Text)
Publication 20020606 A2 Without international search report and to be
                         republished upon receipt of that report.
Search Rpt   20040226 Late publication of international search report
Republication 20040226 A3 With international search report.

Main International Patent Class: **G06F-017/30**
Fulltext Availability:
  Detailed Description

Detailed Description
... browser window of the contents of the document whose summary 62 is
  currently in the **summary** pane 64. In an alternative embodiment,
  clicking on the **document** title results in the **replacement** of the
  contents of the primary window with the contents of the document whose
  summary...

 47/5,K/42   (Item 42 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT
(c) 2005 WIPO/Univentio. All rts. reserv.

00885041   **Image available**
CONCEPT IDENTIFICATION SYSTEM AND METHOD FOR USE IN REDUCING AND/OR
    REPRESENTING TEXT CONTENT OF AN ELECTRONIC DOCUMENT
PROCEDE ET SYSTEME D'IDENTIFICATION DE CONCEPT, UTILES POUR REDUIRE ET/OU
    REPRESENTER UN CONTENU TEXTE D'UN DOCUMENT ELECTRONIQUE
Patent Applicant/Assignee:
  AMIKANOW CORPORATION, Suite 400A, 320 March Road, Kanata, Ontario K2K 2E3
  , CA, CA (Residence), CA (Nationality)
Inventor(s):
  ABU-HAKIMA Suhayya, 22 Balding Crescent, Kanata, Ontario K2K 2L4, CA,
Legal Representative:
  CASSAN Lynn (et al) (agent), Cassan MaClean, Suite 401, 80 Aberdeen St.,
  Ottawa, Ontario K1S 5R5, CA,
Patent and Priority Information (Country, Number, Date):
  Patent: WO 200219155 A2-A3 20020307 (WO 0219155)
  Application: WO 2001CA1197 20010828 (PCT/WO CA0101197)
  Priority Application: US 2000649028 20000828
Designated States:
(Protection type is "patent" unless otherwise stated - for applications
prior to 2004)
  AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ
  EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR
  LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ PH PL PT RO RU SD SE SG SI SK
  SL TJ TM TR TT TZ UA UG UZ VN YU ZA ZW
  (EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE TR

English Abstract
  A concept identification system useful in reducing and/or representing
  text content of an electronic document and in highlighting the content of
  the document. A concept knowledge base comprises a plurality of concepts
  and each concept comprises one or more subconcepts linked to each other
  and to the concept on a hierarchical basis. One or more of the
  subconcepts may be linked to one or more subconcepts of another concept.
  A concept matching module matches text of the document to subconcepts of
  the concept knowledge base and assesses any links between the matched
  subconcepts and other concepts and/or subconcepts of the concept
  knowledge base. From this a determination is made of whether the document
  relates to a concept of the knowledge base. With an identification of
  such concept a document representation generator may produce a precis of
  the document based on a template associated with such concept. For
  highlighting of a document a highlighter module determines key content of
  the input document and an interface integrates the concept identification
  system and the highlighter module. An output module produces an output
  highlight document from the key content.

French Abstract
  L'invention concerne un systeme d'identification de concept, utile pour
  reduire et/ou representer un contenu texte d'un document electronique et
  pour mettre en evidence le contenu de ce document. Une base de
  connaissances de concepts comprend plusieurs concepts comprenant chacun
  un ou plusieurs sous-concepts lies les uns aux autres ainsi qu'au
  concept, sur une base hierarchique. Un ou plusieurs sous-concepts peuvent
  etre lies a un ou plusieurs sous-concepts d'un autre concept. Un module
  de mise en correspondance de concepts met en correspondance le texte du
  document avec les sous-concepts de la base de connaissances de concepts
  et determine tous les liens entre les sous-concepts mis en correspondance
  et d'autres concepts et/ou sous-concepts de la base de connaissances de
  concepts. Apres cette etape de mise en correspondance, une etape de
  determination est executee qui permet de savoir si le document se
  rapporte a un concept de la base de connaissances. Grace a
  l'identification d'un tel concept, un generateur de representation de
  documents peut produire un abrege du document, d'apres un modele associe
  a un tel concept. Pour mettre en valeur un document, un module de mise en
  valeur determine le contenu cle du document entre et une interface
  integre le systeme d'identification de concepts et le module de mise en
  valeur. Un module de sortie produit un document mis en valeur a partir du
  contenu cle.

Republication 20031224 A3 Before the expiration of the time limit for
                         amending the claims and to.be republished in the
                         event of the receipt of amendments.

Main International Patent Class: **G06F-017/30**
Fulltext Availability:
  Detailed Description

Detailed Description
... purposes of
  assisting the user in handling such communications and for
  effective classification, archiving and **retrieval** of the
  information.

    SUBSTITUTE SHEET (RULE 26)
  The known **document   condensers** (sometimes also referred
  to as key word/phrase "extractors" or as "summarizers") ,
  which typically function...network to
  improve the highlighter system's assignment of weightings to
  other words of the **document** for purposes of generating a
  14
    SUBSTITUTE SHEET (RULE 26)
  highlight **summary** of the **document** as detailed in said copending U.S.
  application.

  Identification of a concept by the system...

...conference room 101 between 1:30 and 3:00pm for all managers."
  The system thereby **substitutes** standardized terms for terms of
  the **document** to form a **precis** text that is much clearer for
  the user than would be produced by simply extracting...

 47/5,K/54   (Item 54 from file: 349)
DIALOG(R)File 349:PCT FULLTEXT

00805456   **Image available**
CONTENT-SPECIFIC FILENAME SYSTEMS
SYSTEMES DE NOMS DE FICHIERS SPECIFIQUES AU CONTENU
Patent Applicant/Assignee:
  MICROSOFT CORPORATION, One Microsoft Way, Redmond, WA 98052, US, US
    (Residence), US (Nationality)
Inventor(s):
  BURGESS Giles J, 845 Bellevue Pl. East, Apt. 301, Seattle, WA 98102, US,
Legal Representative:
  PAYNE Stephen S (et al) (agent), Banner & Witcoff, Ltd., 1001 G Street,
    N.W., Eleventh Floor, Washington, DC 20001-4597, US,
Patent and Priority Information (Country, Number, Date):
  Patent: WO 200139043 A2-A3 20010531 (WO 0139043)
  Application: WO 2000US26216 20000925 (PCT/WO US0026216)
  Priority Application: US 99447787 19991123
Designated States:
(Protection type is "patent" unless otherwise stated - for applications
prior to 2004)
  AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CR CU CZ DE DK DM DZ EE
  ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT
  LU LV MA MD MG MK MN MW MX MZ NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM
  TR TT TZ UA UG UZ VN YU ZA ZW
  (EP) AT BE CH CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE

(OA) BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG
(AP) GH GM KE LS MW MZ SD SL SZ TZ UG ZW
(EA) AM AZ BY KG KZ MD RU TJ TM
Main International Patent Class: **G06F-017/30**
Publication Language: English
Filing Language: English
Fulltext Availability:
  Detailed Description
  Claims
Fulltext Word Count: 8361

English Abstract
  A computer file naming technique employs content-specific filenames
  (CSFN's) that represent globally-unique identifiers for the contents of a
  file. Since file references incorporating the CSFN's are not
  location-specific, they offer unique advantages in the areas of file
  caching and file installation. Particularly, web browsers enabled to
  recognize CSFN's inherently verify the content of files when they are
  retrieved from a local cache, eliminating the need for comparison of file
  data or time stamps of the cached file copy and the server copy. Thus,
  file verification occurs solely in the local context. The invention
  includes caching and software installation systems that incorporate the
  benefits of CSFN's.

French Abstract
  Selon l'invention, une technique de denomination d'un fichier
  informatique utilise des noms de fichiers specifiques au contenu (CSFN)
  representant des identificateurs globalement uniques du contenu d'un
  fichier. Etant donne que les references de fichier integrant les noms de
  fichiers specifiques au contenu ne sont pas specifiques a l'emplacement,
  elles presentent des avantages uniques dans les domaines de mise en
  antememoire et d'installation de fichiers. Plus particulierement, les
  navigateurs Web pouvant reconnaitre lesdits noms de fichiers specifiques
  au contenu verifient intrinsequement le contenu des fichiers lorsque ces
  derniers sont recuperes dans une antememoire locale, ce qui rend inutile
  la comparaison des donnees de fichier ou des horodateurs de la copie du
  fichier mise en antememoire et de la copie sur serveur. Ainsi, les
  fichiers ne sont verifies que localement. L'invention concerne des
  systemes de mise en antememoire et d'installation de logiciels integrant
  les avantages des noms de fichiers specifiques au contenu.

Legal Status (Type, Date, Text)
Publication   20010531 A2 Without international search report and to be
                        republished upon receipt of that report.
Examination   20010927 Request for preliminary examination prior to end of
                        19th month from priority date
Search Rpt   20031211 Late publication of international search report
Republication 20031211 A3 With international search report.
Republication 20031211 A3 Before the expiration of the time limit for
                        amending the claims and to be republished in the
                        event of the receipt of amendments.

Main International Patent Class: **G06F-017/30**
Fulltext Availability:
  Detailed Description

Detailed Description
... technique for reducing the duplication of content, i.e., logos,
  backgrounds, bars, buttons, etc., in **retrieved** HTML and text documents
  on the Web.

Using the technique of Mogul and van Hoff, any response whose message **digest** is equivalent to the **message** **digest** of the **requested** resource may be **substituted** . A
3
proxy may check its cache to see if a cached instance of the...

...both the client and server. Moreover, their method requires the additional step of determining the **message** **digest** of the **requested** resource before **substitution** can occur. This additional step prevents back-compatibility of the technique of Mogul and van...
?
?

Set	Items	Description
S1	374125	DISPLAC? OR SUBROGAT?
S2	2299055	REPLAC??? ? OR REPLACEMENT? OR SUBSTITUT???? ? OR RE()(PLAC??? ? OR PLACEMENT? ?) OR EXCHANG? OR SWAP? OR SWITCH? OR SUPPLANT?
S3	417	FULLTEXT? OR (FULL OR COMPLETE)()TEXT
S4	1510714	DOCUMENT? ? OR ARTICLE? ? OR RECORD? ? OR REPORT? ? OR MESSAGE? ?
S5	3454309	ABSTRACT? ? OR SUMMARY? OR SUMMARIES OR SUMMATION? OR SHORTEN? OR SHORTER? OR ABBREVIAT? OR REDUC??? ? OR REDUCTION?
S6	327132	ABRIDG? OR CONDENS??? ? OR PRECIS OR SYNOPSI? OR CAPSUL? OR RECAP? ? OR BRIEF?? ? OR DIGEST? ?
S7	273408	EXTRACT? ?
S8	27855	S1:S2(10N)S3:S4
S9	772	S8(10N)S5:S7
S10	1075188	SEARCH? OR RETRIEV? OR HARVEST? OR QUERY? OR QUERIE? OR MINE? ? OR MINING OR DATAMIN? OR TEXTSEARCH? OR REQUEST?
S11	52509	IR
S12	97	S11(3N)S4
S13	0	S9 AND S12
S14	119	S9 AND S10:S11
S15	153002	DATA()(BASE? ? OR SET? ? OR BANK? ? OR STORE? OR SYSTEM? OR COLLECTION? OR LIBRAR? OR DEPOSIT? OR REPOSIT? OR WAREHOUSE? OR WARE? ? OR MART? ?)
S16	8	DATAWAREHOUSE? OR DATAMART?
S17	110165	DATABASE? OR DATASET? OR DATABANK? OR DATASTORE? OR DATASYSTEM? OR DATACOLLECTION? OR DATALIBRAR? OR DATADEPOSIT? OR DATAREPOSIT?
S18	49	S9 AND S15:S17
S19	1237282	IC='G06F'
S20	45	S14 AND S19
S21	91	S16 OR S18 OR S20
S22	29	S21 AND AC=US/PR
S23	19	S22 AND AY=(1976:2000)/PR
S24	51	S21 AND PY=1976:2000
S25	60	S23:S24
S26	60	IDPAT (sorted in duplicate/non-duplicate order)
S27	58	IDPAT (primary/non-duplicate records only)

27/9/3   (Item 3 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2005 Thomson Derwent. All rts. reserv.

014385686   **Image available**
WPI Acc No: 2002-206389/200226
Related WPI Acc No: 2002-206368
XRPX Acc No: N02-157180
  Content delivery accelerating system for use in networked environment to
  replace **old** documents **uses** condenser **to** replace **unchanged portions**
  of document **by pointers to these portions**
Patent Assignee: FINEGROUND NETWORKS (FINE-N)
Inventor: KAUSIK B N
Number of Countries: 095 Number of Patents: 002

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
WO 200213037	A1	20020214	WO 2001US24936	A	20010808	200226	B
AU 200181205	A	20020218	AU 200181205	A	20010808	200244	

Priority Applications (No Type Date): US 2000634134 A 20000808
Patent Details:

Patent No	Kind Lan Pg	Main IPC	Filing Notes
WO 200213037	A1 E 25	G06F-015/16	

    Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA
    CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN
    IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ
    PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG UZ VN YU ZA ZW
    Designated·States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR
    IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW

AU 200181205 A   G06F-015/16   Based on patent WO 200213037

Abstract (Basic): WO 200213037 A1
    NOVELTY - A condenser (200) ships the assembly module (120) from a
content server (300) as a self-unwrapping Javascript process to the
user computer (100) during the initial reaction and the condenser is
configured transparently. The condenser maintains historic information
about pages most frequently **requested** by each user so that, when a
document is **requested** , only changed portions need to be reproduced.
    DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for a method
of improving network efficiency, for a condenser, for a computer
readable medium with processing instructions and for a method and
system for reassembling a condensed document.
    USE - Accelerating delivery of content in networked environment.
    DESCRIPTION OF DRAWING(S) - The drawing shows the system
    Condenser (200)
    Assembly module (120)
    Content server (300)
    User computer (100)
    pp; 25 DwgNo 1/5

Title Terms: CONTENT; DELIVER; ACCELERATE; SYSTEM; ENVIRONMENT; REPLACE;
  DOCUMENT; CONDENSER; REPLACE; UNCHANGED; PORTION; DOCUMENT; POINT;
  PORTION
Derwent Class: T01
International Patent Class (Main): **G06F-015/16**
File Segment: EPI
Manual Codes (EPI/S-X): T01-D02; T01-F07; T01-H03A; T01-N03A1; T01-S03

27/9/11   (Item 11 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2005 Thomson Derwent. All rts. reserv.

013575192   **Image available**
WPI Acc No: 2001-059399/ **200107**
XRPX Acc No: N01-044290
  **Organizing electronic mail for client with more than one work station,
  comprises modifying mail with large attachments to make it more suitable
  for access by mobile phone**
Patent Assignee: TELEFONAKTIEBOLAGET ERICSSON L M (TELF)
Inventor: AQUILON A; GABRIELSSON J; WILLEHADSON S
Number of Countries: 092 Number of Patents: 005
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
SE 9902462	A	20001106	SE 992462	A	19990628	200107	B
SE 513804	C2	20001106				200107	

Priority Applications (No Type Date): SE 992462 A 19990628
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
SE 9902462	A		24	G06F-017/60	
WO 200101663	A1	E		H04M-003/533	

    Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY CA CH
CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE
KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU
SD SE SG SI SK SL TJ TM TR TT TZ UA UG UZ VN YU ZA ZW
    Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR
IE IT KE LS LU MC MW NL OA PT SD SE SL SZ TZ UG ZW

| AU 200044467 | A | | | H04M-003/533 | Based on patent WO 200101663 |
| US 6738800 | B1 | | | G06F-013/00 | |

Abstract (Basic): SE 9902462 A
    NOVELTY - Electronic information (101) is received in its original
format, stored in a first directory (102) and processed to generate
information (113a-113c) adapted to a different data format, this
processed information then being stored in a directory (121a-121c) with
a structure corresponding to that of the first directory. The client
then **retrieves** the information either in its original or processed
data format, depending on the access environment of his or her work
station.
    DETAILED DESCRIPTION - An INDEPENDENT CLAIM is also included for
the device used to carry out this process.
    USE - None given.
    ADVANTAGE - Large attachments such as images, video clips or sound
files can be removed when email is sent to a client's mobile phone
(e.g. WAP phone), reducing the burden on the mail organizing system.
    DESCRIPTION OF DRAWING(S) - Figure 1 shows a schematic view of the
electronic mailbox.
    Storage area (100)
    Incoming electronic mail (101)
    User-defined directory of stored mail (102)
    Agent (110)
    Analysis device (111)
    Electronic mail processor (112)
    Processed mail (113a-113c)
    Storage area (120)
    Directories corresponding to user-defined directory of stored mail
(121a-121c)
    Ethernet (130)
    Work station (135)
    Radio link (140)
    Mobile phone (145)
    Modem link (150)
    Laptop computer (155)
    Modem link (160)
    Work station (165)
    Interface, e.g. electronic mail server (170)
    pp; 24 DwgNo 1/5
Technology Focus:
    TECHNOLOGY FOCUS - IMAGING AND COMMUNICATION - Processing involves
filtering at least one component of the electronic information in order
to convert it into a new format, especially by removing a colour
component from the information, or by removing at least part of one of
the information components to **reduce** its size and **replacing** with an

identifier. Images and voice **messages** can be included or attached to
the information.
Title Terms: ORGANISE; ELECTRONIC; MAIL; CLIENT; MORE; ONE; WORK; STATION;
  COMPRISE; MODIFIED; MAIL; ATTACH; MORE; SUIT; ACCESS; MOBILE; TELEPHONE
Derwent Class: T01; W01; W02
International Patent Class (Main): **G06F-013/00 ; G06F-017/60 ;**
  H04M-003/533
International Patent Class (Additional): H04L-012/58
File Segment: EPI
Manual Codes (EPI/S-X): T01-E01A; T01-H07C1; T01-J05B4P; W01-A06B7;
  W01-A06E1; W01-A06G2; W01-A06X; W01-B05A1A; W02-C03C1A

27/9/14   (Item 14 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2005 Thomson Derwent. All rts. reserv.

012857182   **Image available**
WPI Acc No: 2000-029015/ **200003**
XRPX Acc No: N00-022039
   Search **substituted processing apparatus for document processing device**
   - searches **document data by replacing corresponding character row data**
   into another character row data
Patent Assignee: CANON KK (CANO)
Number of Countries: 001 Number of Patents: 001
Patent Family:
Patent No   Kind   Date   Applicat No   Kind   Date   Week
JP 11296513   A   19991029   JP 98104591   A   19980415   200003 B

Priority Applications (No Type Date): JP 98104591 A 19980415
Patent Details:
Patent No Kind Lan Pg   Main IPC   Filing Notes
JP 11296513   A   7 G06F-017/24

Abstract (Basic): JP 11296513 A
      NOVELTY - **Search** substitution unit performs **search** substitution
   by replacing the row of characters by another character row.
   automatically. Edit unit judges whether substitution occurs and **search**
   unit proceeds **search** for document data corresponding to the row of
   characters. DETAILED DESCRIPTION - An INDEPENDENT CLAIM is also
   included for procedure for **search** substituted process.
      USE - For document processing.
      ADVANTAGE - The **searched** character row is **replaced** by another
   character row, thus the corresponding **document** data is preserved.
   Therefore operators burden is **reduced** , operation efficiency is hence
   achieved. DESCRIPTION OF DRAWING(S) - The figure shows block diagram of
   document processor.
      Dwg.7/7
Title Terms: **SEARCH** ; SUBSTITUTE; PROCESS; APPARATUS; DOCUMENT; PROCESS;
  DEVICE; **SEARCH** ; DOCUMENT; DATA; REPLACE; CORRESPOND; CHARACTER; ROW;
  DATA; CHARACTER; ROW; DATA
Derwent Class: T01
International Patent Class (Main): **G06F-017/24**
International Patent Class (Additional): **G06F-017/21**
File Segment: EPI
Manual Codes (EPI/S-X): T01-J11A
? t27/9/26,36

27/9/26   (Item 26 from file: 350)
DIALOG(R)File 350:Derwent WPIX

011272343   **Image available**
WPI Acc No: 1997-250246/ **199723**
XRPX Acc No: N97-206675
  **Electronic filing system for document management - uses link information
  memory to store link information on each document with same** search
  **keyword corresponding to classification of each class of
  document-classification directory**
Patent Assignee: RICOH KK (RICO)
Number of Countries: 001   Number of Patents: 001
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
JP 9081585	A	19970328	JP 95236466	A	19950914	199723	B

Priority Applications (No Type Date): JP 95236466 A 19950914
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
JP 9081585	A		17	G06F-017/30	

Abstract (Basic): JP 9081585 A
      The electronic filing system has an adder which adds the keyword
  for **searching** , and the classification contents are displayed. A
  document memory (206) stores the classification contents and the
  keyword for **searching** , corresponding to the specific document. A
  selector chooses some classes of a document-classification directory
  from the contents of the document memory. An information memory stores
  the document- management information related to every user,
  corresponding to the hierarchical classification of the
  document-classification directory.
      A **searching** part looks for the document with the same **search**
  keyword obtained from the document memory, based on the classification
  of each class of document-management information. A link memory stores
  the link information on each document with the same **search** keyword,
  corresponding to the classification of each class of the
  document-classification directory.
      ADVANTAGE - Enables automatic deletion and **substitution**
  corresponding to **document** -management information. **Reduces   document**
  -management burden on user. Avoids causing damage to original image of
  registered document.
      Dwg.1/11
Title Terms: ELECTRONIC; FILE; SYSTEM; DOCUMENT; MANAGEMENT; LINK;
  INFORMATION; MEMORY; STORAGE; LINK; INFORMATION; DOCUMENT; **SEARCH** ;
  KEYWORD; CORRESPOND; CLASSIFY; CLASS; DOCUMENT; CLASSIFY; DIRECTORY
Derwent Class: T01
International Patent Class (Main): **G06F-017/30**
International Patent Class (Additional): **G06F-012/00** ; **G06F-017/21**
File Segment: EPI
Manual Codes (EPI/S-X): T01-J05B1; T01-J05B3; T01-J11D

 27/9/36   **(Item 36 from file: 350)**
DIALOG(R)File 350:Derwent WPIX

008690738
WPI Acc No: 1991-194758/ **199127**
XRPX Acc No: N91-149124
  Image forming-storing apparatus e.g. copying machine - replaces **sheets
  of** document **data with one sheet of paper on which** abstract **image data**

is recorded
Patent Assignee: TOSHIBA KK (TOKE)
Inventor: HASEGAWA H; MAEDA M; MIURA K; NAKAMURA H
Number of Countries: 003 Number of Patents: 002
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
EP 435174	A	19910703	EP 90124906	A	19901220	199127	B
EP 435174	A3	19920415	EP 90124956	A	19901220	199328	

Priority Applications (No Type Date): JP 89336008 A 19891225
Cited Patents: NoSR.Pub; EP 327931; EP 330343; GB 2219674
Patent Details:
Patent No Kind Lan Pg Main IPC   Filing Notes
EP 435174   A
   Designated States (Regional): DE FR GB
Abstract (Basic): EP 435174 A
      Image data is converted into digital data and is read by a scanner
   (41). The digital data read by the scanner (41) is stored in an optical
   disk (233). **Retrieval** data (802,805) representing the storage
   position or the like of the **data   stored** in the optical disk (233)
   is also recorded in the optical disk (233). In addition, paper (P) on
   which abstract image data of the digital **data   stored** in the optical
   disk (233) is recorded is output from a printer (43). The respective
   components of the invention are integrally arranged. Sheets of
   document data are **replaced** with one sheet of paper (P) on which
   simple **abstract** image data is recorded and image **data   stored** in
   an optical disk (233), thus allowing easy registration and **retrieval**
   of image data of documents and enabling a great reduction in space for
   documents.
      ADVANTAGE - Enjoys both advantages of paper file and electronic
   file, i.e., quick data access and large storage capacity. (22pp
   Dwg.No.1/10
Title Terms: IMAGE; FORMING; STORAGE; APPARATUS; COPY; MACHINE; REPLACE;
  SHEET; DOCUMENT; DATA; ONE; SHEET; PAPER; ABSTRACT; IMAGE; DATA; RECORD
Derwent Class: S06; T01; W02; W04
International Patent Class (Additional): **G06F-015/64**
File Segment: EPI
Manual Codes (EPI/S-X): S06-A16; T01-H01B; T01-J05B; T01-J10A; W02-J09;
  W04-K05
?
PLEASE ENTER A COMMAND OR BE LOGGED OFF IN 5 MINUTES
? t27/9/39,41-42,55-56,58

 27/9/39   (Item 39 from file: 347)
DIALOG(R)File 347:JAPIO
(c) 2005 JPO & JAPIO. All rts. reserv.

06271340   **Image available**
DATA PROCESSING DEVICE AND METHOD, DOCUMENT EDITING SYSTEM AND RECORD
MEDIUM

PUB. NO.:   11-212928 [JP 11212928 A]
PUBLISHED: August 06, 1999 (**19990806)**
INVENTOR(s): ITO KOICHI
              SEKINE MINORU
              BANDO HIROYUKI
APPLICANT(s): SONY CORP
APPL. NO.:   10-013903 [JP 9813903]
FILED: January 27, 1998 (19980127)
INTL CLASS: **G06F-015/16**

ABSTRACT

PROBLEM TO BE SOLVED: To quickly edit or print the document data.

SOLUTION: When the document data are inputted by a document data input means 11 and the image data are interpolated to a document, a client 1 gives a transfer **request** to an image server 2 via an editing image data transfer **request** means 12. Receiving the transfer **request** from the client 1, the server 2 reads out the editing image data recorded by an editing image data recording means 24 and transmits them to the client 1. The client 1 receives the editing image data, interpolates them to the document data and sends them to a data processor 3. The processor 3 **extracts** URL from the editing image data interpolated to the **document**, acquires the corresponding image data from the server 2 to **replace** them with the editing image data and prints these editing image data by a printer, etc.

27/9/41   (Item 41 from file: 347)

05822271   **Image available**
DEVICE AND METHOD FOR READING DOCUMENT ALOUD

PUB. NO.:   10-105371 [JP 10105371 A]
PUBLISHED: April 24, 1998 (**19980424)**
INVENTOR(s): OTANI NORIKO
              IKEDA YUJI
              FUJITA MINORU
APPLICANT(s): CANON INC [000100] (A Japanese Company or Corporation), JP
              (Japan)
APPL. NO.:   08-260841 [JP 96260841]
FILED: October 01, 1996 (19961001)
INTL CLASS: [6] **G06F-003/16 ;   G06F-017/21** ; H04M-003/42
JAPIO CLASS: 45.3 (INFORMATION PROCESSING -- Input Output Units); 36.4
              (LABOR SAVING DEVICES -- Service Automation); 44.4
              (COMMUNICATION -- Telephone); 45.4 (INFORMATION PROCESSING --
              Computer Applications)
JAPIO KEYWORD:R108 (INFORMATION PROCESSING -- Speech Recognition &
              Synthesis); R131 (INFORMATION PROCESSING -- Microcomputers &
              Microprocessers)

ABSTRACT
PROBLEM TO BE SOLVED: To make it possible to grasp the whole contents of an unread document in a short time through the document reading-aloud device which converts an electronized document into a synthesized voice and outputs it in response to an instruction passed through a telephone line.

SOLUTION: An unread document **retrieval** part 102 **retrieves** unread documents from documents held in a document holding part 101 and then the obtained number of the unread documents is reported to a user by an unread document quantity transmission part 104. Further, read-out contents inputted by a read-aloud contents instruction part 105 are referred to and a summary sentence generation part 108 or unread document extraction part 109 generates read-aloud sentences to be voiced out; and a speech synthesis part 111 performs a speech synthesizing process by referring to the read-aloud sentences and obtained speech parameters are outputted by a

speech output part 113, thereby **switching** a read between the **summary** sentences and the whole **document** .

27/9/42 **(Item 42 from file: 347)**
DIALOG(R)File 347:JAPIO
(c) 2005 JPO & JAPIO. All rts. reserv.

04848758 **Image available**
DOCUMENT PROCESSOR

PUB. NO.: 07-141358 [JP 7141358 A]
PUBLISHED: June 02, 1995 (**19950602)**
INVENTOR(s): SOMA TSUNENORI
APPLICANT(s): CANON INC [000100] (A Japanese Company or Corporation), JP
(Japan)
APPL. NO.: 05-288315 [JP 93288315]
FILED: November 17, 1993 (19931117)
INTL CLASS: [6] G06F-017/24
JAPIO CLASS: 45.4 (INFORMATION PROCESSING -- Computer Applications)
JAPIO KEYWORD:R004 (PLASMA); R011 (LIQUID CRYSTALS); R131 (INFORMATION
PROCESSING -- Microcomputers & Microprocessers)

ABSTRACT
PURPOSE: To **shorten** the operation time and to efficiently perform the **full** - **text** **exchange** processing by providing a control means which always executed **full** - **text** **exchange** in accordance with contents of a memory for **full** - **text** **exchange** independently of contents of a **document** memory and performing **full** - **text** **exchange** with the same contents without setting contents to be subjected to full-text exchange with respect to each read document.

CONSTITUTION: On a display part 2, required information and the result after execution are displayed in a window displayed in a prescribed position at a prescribed timing in accordance with the operation on a keyboard 1. A cursor is displayed on this display part 2 to point the input start position of document information or the like, and character information or the like from the keyboard 1 is displayed. A CPU 3 executes various information processings based on control programs and **data** **stored** in a ROM/RAM part 4. Full-text exchange is always executed in accordance with a memory 7 for full-text exchange independently of contents of a · document memory 6. That is, deletion of contents of the memory 7 for full-text exchange is inhibited.

27/9/55 **(Item 55 from file: 347)**
DIALOG(R)File 347:JAPIO
(c) 2005 JPO & JAPIO. All rts. reserv.

02814762 **Image available**
DOCUMENT PRODUCING DEVICE

PUB. NO.: 01-112362 [JP 1112362 A]
PUBLISHED: May 01, 1989 (**19890501)**
INVENTOR(s): KINUGAWA YUKIE
MIMURA YOSHISUKE
APPLICANT(s): MATSUSHITA ELECTRIC IND CO LTD [000582] (A Japanese Company
or Corporation), JP (Japan)
APPL. NO.: 62-269651 [JP 87269651]
FILED: October 26, 1987 (19871026)

INTL CLASS: [4] G06F-015/20
JAPIO CLASS: 45.4 (INFORMATION PROCESSING -- Computer Applications)
JAPIO KEYWORD:R139 (INFORMATION PROCESSING -- Word Processors)
JOURNAL: Section: P, Section No. 913, Vol. 13, No. 353, Pg. 53, August
               08, 1989 (19890808)

ABSTRACT

PURPOSE: To efficiently produce a desired **document** by previously setting
and storing **switch** representing information and **reducing** an editing
operation by the use of the switch representing information.

CONSTITUTION: A switch representing information setting part 1 sets the
switch representing information and a switch representing information
temporary storing part 2 stores the switch representing information. When
an operator desired to use the switch representing information stored in
the switch representing information temporary storing part 2, a switch
executing part 5 switches a part of an edited document to the switch
representing information taken out from the switch representing information
temporary storing part 2. Thereby, the burden of the operator can be
reduced since the operator has to store plural representations and a
document can be formed in a short time since it is not required to reread
all sentence and **search** a deleted part when a mark is set to a part to be
saved previously if the document limiting the number of characters is
formed.

27/9/56   (Item 56 from file: 347)
DIALOG(R)File 347:JAPIO
(c) 2005 JPO & JAPIO. All rts. reserv.

02598286   **Image available**
MIXED MODE DOCUMENT **RETRIEVING** DEVICE

PUB. NO.:   63-215186 [JP 63215186 A]
PUBLISHED: September 07, 1988 (**19880907)**
INVENTOR(s):   ASABA SHOJI
APPLICANT(s):   MATSUSHITA GRAPHIC COMMUN SYST INC [330729] (A Japanese
                Company or Corporation), JP (Japan)
APPL. NO.:   62-048162 [JP 8748162]
FILED: March 03, 1987 (19870303)
INTL CLASS: [4] H04N-007/173; **G06F-015/40** ; H04N-007/12
JAPIO CLASS: 44.6 (COMMUNICATION -- Television); 45.4 (INFORMATION
                PROCESSING -- Computer Applications)
JOURNAL: Section: E, Section No. 701, Vol. 13, No. 5, Pg. 8, January
                09, 1989 (19890109)

ABSTRACT
PURPOSE: To remarkably **shorten** the transmission time of a **document**
trially **retrieved** in order to seach an odjective **document** by
replacing image data in a **retrieved** mixed **document** with pseudo image
data in accordance with a **request** from a terminal equipment.

CONSTITUTION: A **retrieving** device 10 replaces only an image data part
with large data capacity out of a **retrieved** mixed document by the pseudo
image data expressing a fully white image e.g. in accordance with a
request from a terminal equipment. Namely, a document consisting of
character data and image data is converted into a document consisting of
the character data and pseudo image data. In the converted document, the
character part is stored as it is, but the image part is lost and blanked.
Since the pseudo image data express only the blank, the data volume can be

remarkably reduced as compared to that of real image data expressing the original image. Thus, the time required for the transmission can be remarkably shortened.

27/9/58   (Item 58 from file: 347)
DIALOG(R)File 347:JAPIO
(c) 2005 JPO & JAPIO. All rts. reserv.

01624961   **Image available**
SUBSTITUTE KEY CONTROL PROCESSING SYSTEM USING PLURAL ITEMS

PUB. NO.:   60-103461 [JP 60103461 A]
PUBLISHED: June 07, 1985 (**19850607)**
INVENTOR(s):   HONDA SHIGEO
APPLICANT(s): FUJITSU LTD [000522] (A Japanese Company or Corporation), JP
               (Japan)
APPL. NO.:   58-211273 [JP 83211273]
FILED: November 10, 1983 (19831110)
INTL CLASS: [4] **G06F-012/00**
JAPIO CLASS: 45.2 (INFORMATION PROCESSING -- Memory Units)
JOURNAL: Section: P, Section No. 395, Vol. 09, No. 253, Pg. 59,
               October 11, 1985 (19851011)

ABSTRACT
PURPOSE: To evade the rearrangement of a record item and the duplication of this item by defining substitue keys for optional plural items within a desired record.

CONSTITUTION: A substitute key control information memory area 12 stores the information on the compounded items in case a substitute key consists of said items. For instance, the area 12 stores the number of discontinuous items included in the substitute key, the position information, the length, etc. A substitude key defining part 13 refers to each **record** of a master file 10 for **substitute** keys of designated items. Then the part 13 **extracts** all **records** containing the contents of the **substitute** key for each said contents and sets the pointer information to produce a substitute index file 11. When plural items are designated, the record is **retrieved** by the contents combining these plural imtes. Then the control information is stored to the area 12 for automatic renewal of the file 11 when the corresponding record is extracted or the record is added to or deleted from the file 10.
?

```
File    9:Business & Industry(R) Jul/1994-2005/Jul 29
            (c) 2005  The Gale Group
File   16:Gale Group PROMT(R) 1990-2005/Jul 28
            (c) 2005 The Gale Group
File   47:Gale Group Magazine DB(TM) 1959-2005/Aug 01
            (c) 2005 The Gale group
File  148:Gale Group Trade & Industry DB 1976-2005/Jul 29
            (c)2005 The Gale Group
File  160:Gale Group PROMT(R) 1972-1989
            (c) 1999 The Gale Group
File  275:Gale Group Computer DB(TM) 1983-2005/Aug 01
            (c) 2005 The Gale Group
File  570:Gale Group MARS(R) 1984-2005/Jul 29
            (c) 2005 The Gale Group
File  621:Gale Group New Prod.Annou.(R) 1985-2005/Aug 01
            (c) 2005 The Gale Group
File  636:Gale Group Newsletter DB(TM) 1987-2005/Jul 29
            (c) 2005 The Gale Group
File  649:Gale Group Newswire ASAP(TM) 2005/Jul 18
            (c) 2005 The Gale Group


Set      Items    Description
S1      145381    DISPLAC? OR SUBROGAT?
S2     8430118    REPLAC??? ? OR REPLACEMENT? OR SUBSTITUT???? ? OR RE()(PLA-
                  C??? ? OR PLACEMENT? ?) OR EXCHANG? OR SWAP? OR SWITCH? OR SU-
                  PPLANT?
S3      488713    FULLTEXT? OR (FULL OR COMPLETE)()TEXT
S4    16814819    DOCUMENT? ? OR ARTICLE? ? OR RECORD? ? OR REPORT? ? OR MES-
                  SAGE? ?
S5     6198492    ABSTRACT? ? OR SUMMARY? OR SUMMARIES OR SUMMATION? OR SHOR-
                  TEN? OR SHORTER? OR ABBREVIAT? OR REDUC??? ? OR REDUCTION?
S6     6731414    ABRIDG? OR CONDENS??? ? OR PRECIS OR SYNOPSI? OR CAPSUL? OR
                  RECAP? ? OR BRIEF?? ? OR DIGEST? ?
S7      217682    EXTRACT? ?
S8      953305    S1:S2(10N)S3:S4
S9       56298    S8(10N)S5:S7
S10    6040559    SEARCH? OR RETRIEV? OR HARVEST? OR QUERY? OR QUERIE? OR MI-
                  NE? ? OR MINING OR DATAMIN? OR TEXTSEARCH? OR REQUEST?
S11     149944    IR
S12       1277    S11(3N)S4
S13          1    S9(S)S12
S14        720    S9(S)S10:S11
S15     639965    DATA()(BASE? ? OR SET? ? OR BANK? ? OR STORE? OR SYSTEM? ?
                  OR COLLECTION? OR LIBRAR? OR DEPOSIT? OR REPOSIT? OR WAREHOUS-
                  E? OR WARE? ? OR MART? ?)
S16       4921    DATAWAREHOUS? OR DATAMART?
S17    1995945    DATABASE? OR DATASET? OR DATABANK? OR DATASTORE? OR DATASY-
                  STEM? OR DATACOLLECTION? OR DATALIBRAR? OR DATADEPOSIT? OR DA-
                  TAREPOSIT?
S18        575    S9(S)S15:S17
S19     490307    S1:S2(5N)S3:S4
S20      31190    S19(5N)S5:S7
S21        275    S20(S)S10:S11
S22        252    S20(S)S15:S17
S23         32    S21 AND S22
S24          5    S23/2001:2005
S25         27    S23 NOT S24
S26         17    RD (unique items)
S27    2674201    REPLAC??? ? OR REPLACEMENT? OR SUBSTITUT? OR RE()(PLAC??? ?
                  OR PLACEMENT? ?) OR SWAP? OR SUPPLANT?
```

```
S28    41109   (S1 OR S27)(5N)S3:S4
S29     7321   S28(5N)S5:S7
S30      106   S29(S)S10:S11
S31       44   S29(S)S15:S17
S32      143   S30:S31
S33       56   S32/2001:2005
S34       80   S32 NOT (S33 OR S23 OR S13)
S35       50   RD (unique items)
```

35/3,K/2   (Item 2 from file: 16)
DIALOG(R)File 16:Gale Group PROMT(R)
(c) 2005 The Gale Group. All rts. reserv.

07843477   Supplier Number: 65486654   (USE FORMAT 7 FOR FULLTEXT)
**ETL provides the keys to keeping data relevant, available - Data
  extraction, transformation, and loading fuel the changing information
  needs of e-business applications.(Technology Information)**
Steinacher, Scott
InfoWorld, v22, n39, p74
Sept 25, 2000
Language: English   Record Type: Fulltext Abstract
Document Type: Magazine/Journal; Trade
Word Count: 1295

...   is preferable because it reduces response times and enhances ease
of use.
       Before loading a **data   mart** , programmers typically aggregate
data. Aggregation routines **replace** numerous detail **records** with
relatively few **summary** records. For example, suppose that a year's worth
of sales data is stored in several thousand records in a normalized
database .
       Through aggregation, this data is transformed into fewer summary
records that will be written to...

 35/3,K/43   (Item 3 from file: 275)
DIALOG(R)File 275:Gale Group Computer DB(TM)
(c) 2005 The Gale Group. All rts. reserv.

01452976   SUPPLIER NUMBER: 11321277   (USE FORMAT 7 OR 9 FOR FULL TEXT)
**A DataBase Publisher. (Ventura Software Inc.'s DataBase Publisher report
  generator) (Forum)(Brief Article) (product announcement)**
Antonoff, Michael
PC Sources, v2, n10, p87(1)
Oct, 1991
DOCUMENT TYPE: product announcement   ISSN: 1052-6579   LANGUAGE:
  ENGLISH   RECORD TYPE: FULLTEXT
WORD COUNT: 187   LINE COUNT: 00015

       DataBas e Publisher also provides eight types of dictionaries,
including a substitution dictionary that **replaces   abbreviations** used in
 records with formal names for reports, and an exceptions dictionary that
treats certain document styles, such...
?

```
File    6:NTIS 1964-2005/Jul W4
           (c) 2005 NTIS, Intl Cpyrght All Rights Res
File    2:INSPEC 1969-2005/Jul W4
           (c) 2005 Institution of Electrical Engineers
File    8:Ei Compendex(R) 1970-2005/Jul W3
           (c) 2005 Elsevier Eng.  Info. Inc.
File   57:Electronics & Communications Abstracts 1966-2005/Jul
           (c) 2005 CSA.
File   34:SciSearch(R) Cited Ref Sci 1990-2005/Jul W4
           (c) 2005 Inst for Sci Info
File   56:Computer and Information Systems Abstracts 1966-2005/Jul
           (c) 2005 CSA.
File   35:Dissertation Abs Online 1861-2005/Jul
           (c) 2005 ProQuest Info&Learning
File   60:ANTE: Abstracts in New Tech & Engineer 1966-2005/Jul
           (c) 2005 CSA.
File   65:Inside Conferences 1993-2005/Jul W5
           (c) 2005 BLDSC all rts. reserv.
File   94:JICST-EPlus 1985-2005/Jun W2
           (c)2005 Japan Science and Tech Corp(JST)
File   95:TEME-Technology & Management 1989-2005/Jun W3
           (c) 2005 FIZ TECHNIK
File   99:Wilson Appl. Sci & Tech Abs 1983-2005/Jun
           (c) 2005 The HW Wilson Co.
File  111:TGG Natl.Newspaper Index(SM) 1979-2005/Jul 29
           (c) 2005 The Gale Group
File  144:Pascal 1973-2005/Jul W4
           (c) 2005 INIST/CNRS
File  256:TecInfoSource 82-2005/Jun
           (c) 2005 Info.Sources Inc
File  434:SciSearch(R) Cited Ref Sci 1974-1989/Dec
           (c) 1998 Inst for Sci Info
File  438:Library Lit. & Info. Science 1984-2005/Jun
           (c) 2005 The HW Wilson Co
File  483:Newspaper Abs Daily 1986-2005/Jul 30
           (c) 2005 ProQuest Info&Learning
File  583:Gale Group Globalbase(TM) 1986-2002/Dec 13
           (c) 2002 The Gale Group
```

```
Set     Items    Description
S1      547398   DISPLAC? OR SUBROGAT?
S2      3838798  REPLAC??? ? OR REPLACEMENT? OR SUBSTITUT???? ? OR RE()(PLA-
                 C??? ? OR PLACEMENT? ?) OR EXCHANG? OR SWAP? OR SWITCH? OR SU-
                 PPLANT?
S3       14647   FULLTEXT? OR (FULL OR COMPLETE)()TEXT
S4      7043178  DOCUMENT? ? OR ARTICLE? ? OR RECORD? ? OR REPORT? ? OR MES-
                 SAGE? ?
S5      8574762  ABSTRACT? ? OR SUMMARY? OR SUMMARIES OR SUMMATION? OR SHOR-
                 TEN? OR SHORTER? OR ABBREVIAT? OR REDUC??? ? OR REDUCTION?
S6      1579821  ABRIDG? OR CONDENS??? ? OR PRECIS OR SYNOPSI? OR CAPSUL? OR
                 RECAP? ? OR BRIEF?? ? OR DIGEST? ?
S7       386750  EXTRACT? ?
S8       76163   S1:S2(10N)S3:S4
S9        2359   S8(10N)S5:S7
S10     1969268  SEARCH? OR RETRIEV? OR HARVEST? OR QUERY? OR QUERIE? OR MI-
                 NE? ? OR MINING OR DATAMIN? OR TEXTSEARCH? OR REQUEST?
S11      608698  IR
S12       1720   S11(3N)S4
S13          1   S9 AND S12
S14        177   S9 AND S10:S11
S15        171   S9 AND S10
```

```
S16    552607   DATA()(BASE? ? OR SET? ? OR BANK? ? OR STORE? OR SYSTEM? ?
                 OR COLLECTION? OR LIBRAR? OR DEPOSIT? OR REPOSIT? OR WAREHOUS-
                 E? OR WARE? ? OR MART? ?)
S17       302   DATAWAREHOUS? OR DATAMART?
S18    718793   DATABASE? OR DATASET? OR DATABANK? OR DATASTORE? OR DATASY-
                 STEM? OR DATACOLLECTION? OR DATALIBRAR? OR DATADEPOSIT? OR DA-
                 TAREPOSIT?
S19       143   S9 AND S16:S18
S20     50814   S1:S2(5N)S3:S4
S21       846   S20(5N)S5:S7
S22        76   S21 AND S10:S11
S23        55   S21 AND S16:S18
S24       122   S22:S23 NOT S13
S25        23   S24/2001:2005
S26        99   S24 NOT S25
S27        67   RD (unique items)
? t27/7/55,57,67
```

27/7/55   (Item 16 from file: 8)
DIALOG(R)File   8:Ei Compendex(R)
(c) 2005 Elsevier Eng.   Info. Inc. All rts. reserv.

01042484   E.I. Monthly No: EI8109072818   E.I. Yearly No: EI81024036
  Title: SUMMARY TIME ORIENTED RECORD (STOR).
  Author: O'Keefe, Q. E. Whiting; Simborg, Donald W.
  Corporate Source: Univ of Calif, San Francisco
  Source: Proc Annu Symp Comput Appl Med Care 4th, Proc of the Annu Conf of
the Soc for Adv Med Syst, 12th, vol 2, Washington, DC, Nov 1-5 1980. Publ
by IEEE (Cat n 80CH1570-1), Piscataway, NJ, 1980 p 1175-1182
  Publication Year: 1980
  CODEN: PCMCDC
  Language: ENGLISH
  Journal Announcement: 8109
  Abstract: STOR is computerized three component, time-oriented, **summary** ,
medical **record** designed to partially **replace** the traditional paper
chart in the outpatient clinics at the University of California San
Franciso. Information from at least four distributed **databases**
functioning independently is brought together in a single paper document.
STOR is prioritized, displays inter-problem and chronological
relationships, provides a high degree of physician control over the display
and provides a great deal of information with little manual physician
effort. Besides the usual issues of cost, impact, and acceptance, the
evaluation will address the question of the informational competence of
STOR in two single blind randomized controlled trials. 19 refs.

27/7/57   (Item 18 from file: 8)
DIALOG(R)File   8:Ei Compendex(R)
(c) 2005 Elsevier Eng.   Info. Inc. All rts. reserv.

00277984   E.I. Monthly No: EI7213020823
   Title: **UTILIZATION OF TERSE CONCLUSIONS IN AN INDUSTRIAL RESEARCH**
ENVIRONMENT.
  Author: Gordon, Irving; Carr, Russell L. K.; Bernier, Charles L.
  Corporate Source: Hooker Chemical Corp, Niagara Falls, NY
  Source: Journal of Chemical Documentation v 12 n 2 May 1972 p 86-88
  Publication Year: 1972
  CODEN: JCHDAN
  Language: ENGLISH
  Journal Announcement: 7213
  Abstract: Terse Conclusions are used at Hooker Research Center by

technical and management personnel as concise report surrogates, as an
internal awareness mechanism, and as the key components in a Report Header
Sheet System. A Terse Conclusion is concerned with the meaning of a
document rather than the subject scope, as in an abstract. A Terse
Conclusion is written for every project in a heterogeneous report. Weekly
compendiums of Terse Conclusions, which have **replaced   abstracts** at
Hooker for Progress **Reports** , are circulated to the entire staff. These
Terse Conclusions are also intended to aid in management decisions. 9 refs.

27/7/67   (Item 4 from file: 256)
DIALOG(R)File 256:TecInfoSource
(c) 2005 Info.Sources Inc. All rts. reserv.

00115858   DOCUMENT TYPE: Review

**PRODUCT NAMES: Company--Cambridge Scientific Abstracts (CSA)   (865087);
Company--Academic Press   (867357)**

TITLE: From Cambridge abstract to Academic full text
AUTHOR: Hamilton, Feona J
SOURCE: Information World Review,   v145 p10(1) Mar 1999
ISSN: 0950-9879
HOMEPAGE: http://www.iwr.co.uk

RECORD TYPE: Review
REVIEW TYPE: Company

Cambridge Scientific Abstracts (CSA) and Academic Press (AP) have teamed up
to provide an innovative approach to hypertext linking of scientific
abstracts with full-text article versions in one environment. AP abstracts
will now be made available directly through CSA's own abstract service to
augment the latter's ability to consistently move users from abstracts to
full-text versions. With over 80 Internet **Database** Service **databases**
included now, users are more assured of viewing more consistent information
when **switching** between **databases** , **abstracts** , and **full - text**
versions from disparate sources. Though both CSA and AP already offer
expansive abstract services, the partnership promises to provide a wealthy
resource that is better than the singular sum of its parts. More than 2,000
articles are added every day to AP's resources for the company's over 7
million authorized users in 11 countries. Corporate libraries can now
purchase three-year licenses from AP and CSA.

Set   Items   Description
S1   1   AU=PROKOPH A?
? t1/9

 1/9/1   (Item 1 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2005 Thomson Derwent. All rts. reserv.

014821720   **Image available**
WPI Acc No: 2002-642426/200269
XRPX Acc No: N02-507781
  **Information retrieval method in computer system, involves accessing
  search index storing tokens obtained by decomposing webpage extract to
  retrieve information from webpage extracts that satisfy search query**
Patent Assignee: PROKOPH A (PROK-I)
Inventor: **PROKOPH A**
Number of Countries: 001 Number of Patents: 001
Patent Family:
Patent No   Kind   Date   Applicat No   Kind   Date   Week
US 20020091671 A1   20020711 US 2001989970   A   20011120. 200269 B

Priority Applications (No Type Date): EP 2000125608 A 20001123
Patent Details:
Patent No Kind Lan Pg   Main IPC   Filing Notes
US 20020091671 A1   11 G06F-007/00

Abstract (Basic): US 20020091671 A1
        NOVELTY - A webpage to be indexed, is retrieved and a webpage
  extract is generated. The webpage extract is decomposed into tokens
  which are stored in a search index (207). The search index is accessed
  by a search engine to retrieve information from webpage extracts which
  satisfy a search query.
        DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for the
  following:
        (1) Computer readable medium storing information retrieving
  program; and
        (2) Information retrieval system.
        USE - For information retrieval in computer system.
        ADVANTAGE - Information is retrieved effectively within short time.
  The quality of search result is improved since the retrieved documents
  re more relevant in view of the semantic concept or notion represented
  by search query.
        DESCRIPTION OF DRAWING(S) - The figure shows a block diagram of the
  search engine.
        Search index (207)
        pp; 11 DwgNo 2/3
Title Terms: INFORMATION; RETRIEVAL; METHOD; COMPUTER; SYSTEM; ACCESS;
  SEARCH; INDEX; STORAGE; TOKEN; OBTAIN; DECOMPOSE; EXTRACT; RETRIEVAL;
  INFORMATION; EXTRACT; SATISFY; SEARCH; QUERY
Derwent Class: T01

International Patent Class (Main): G06F-007/00
File Segment: EPI
Manual Codes (EPI/S-X): T01-N03A2; T01-S03
?

INDEXING AND ABSTRACTING

INTRODUCTORY LECTURE

Often, people looking for information cannot identify the specific data or combination of data that they want at the beginning of their searches. It may be that their needs are not yet fully defined, and they simply want to browse. Alternatively, they may not understand a retrieval system sufficiently, or it may simply be impossible for their needs to be expressed in its terms. As a result they will not be able to translate their needs into the appropriate specifications and must scan to find what they want.

At that point the sheer number of information sources is a major problem in retrieving the information desired. Retrieval problems for a small file are relatively trivial. For example, the owner of a small collection of reprints knows the content of each publication. To retrieve a single item one needs only to leaf through the pile of reprints for the needed paper. One could scan the entire collection in order to locate specific pieces of information.

Locating a single item in a large collection is a problem of a different magnitude. One needs to know either the exact location of the item or the general location of items which are on the same topic. A large array of aids has been designed to facilitate the retrieval of specific documents or groups of documents with common subjects. An index is by far the most common adjunct to any sizable collection of documents. Index cards or index records serve as surrogates to the actual documents and may be easily arranged in various ways.

Information Representation

Information representation is that aspect of information retrieval in which the original file of documents is represented by a set of tags or surrogates such as abstracts or index terms. The concept of subject retrieval is also known as content representation. The physical forms of representation are organized in such a manner that they may be manipulated and searched to access more efficiently and effectively the content of the collection. The key concepts are organization of the information resources in order that searching can be facilitated. The aim of organization is not for the sake of organizing. Organization is for the express purpose of expediting information retrieval. Thus, indexing and abstracting is a vital component in the communication link between the originator of information and its ultimate consumer.

Abstracts and indexes organize the literature so that a specialist can identify documents of interest more easily. This is particularly important in scientific and technical fields of endeavor, but it is also becoming increasingly recognized as essential in the social sciences and humanities.

Some of you may find yourselves employed in a library or commercial setting doing indexing and/or abstracting. Others of you will be consumers of the products of indexing and abstracting services. An appreciation of the decisions necessary in the compilation of abstracts and indexes is essential not only to the intending indexer but also to the information professional devoted to information work in whatever setting.

ABSTRACTS: THE BASICS

Definition: The American National Standards Institute (1979) defines an abstract as an abbreviated, accurate representation of the contents of a document, preferably prepared by its author for publication with it. In short, it is a concise condensation of the significant content of a document presenting its objectives, scope, and major findings. Its primary objective is to capture the essential content of the document thus saving the reader's time. Thus, instead of scanning the entire document, the reader may decide on its relevance by reading a short representation of it. An abstract assists the reader in determining whether there is a need to consult the full text in order to gain the needed information. An abstract also contains terms, called index terms, relating to the subject of the document. Thus, the abstract is often an integral part of a bibliographic record in an indexing system that enhances retrievability of the original document.

Good abstracts are highly structured, concise, and coherent, and are the result of a thorough analysis of the content of the abstracted materials. The art of abstracting demands the application of extensive reading, thinking, writing, and editing skills.

Conciseness and significance are two key concepts in abstracting. Both are relative terms, subject to interpretation. Abstractors attempt to write in a clear, terse, accurate, and noncritical manner in a style similar to that of the original publication. Yet, different abstracts may be written for the identical document depending on the audience for whom the abstracts are written. Abstracts from foreign journal articles generally contain more detailed information than those journals are more easily accessible. Most abstracting guidelines and published criteria suggest the inclusion of the objective of the document, the method used, the results, and the conclusion. These distinct components are not necessarily present in all documents.

Types of Abstracts
There are two main types of abstracts found in commercial abstracting services, informative abstracts and indicative or descriptive abstracts. We will also briefly look at a third type the critical abstract or review. The intended use of the original documents often determines the type of abstract written.

1. The informative abstract acts as a substitute for the document. It is a miniature version of the document including the purpose, numerical data, methodologies, formula, conclusions, and recommendations. It is often used for experimental work, and for specific research reports. It presents what has been done. Many abstracting services permit 100 to 500 words for each abstract. [The average is about 250 words.] Writing informative abstracts for reviews and discursive papers on broader subjects is more difficult for many such papers present too many individual and disjoint ideas in the space of a single paper.
2. An indicative abstract describes what a document is about. It does not report on the actual findings. Therefore, it is well suited to state-of-the-art reviews, literary criticism, lengthy texts, descriptive works, and general discussions of a topic. It tends to be shorter than an informative abstract, containing 50 to 100 words. It gives little detail and contains less content than the original document. Indicative abstracts abound in phrases such as "is discussed" or "has been investigated." Since the treatment is more superficial than in an informative abstract, in most cases, an indicative abstract can be written much faster and is less costly to produce than informative abstracts. An indicative abstract is seldom used as a replacement for the original document. Ideally indicative abstracts give the reader ample information as to whether the original document should be read and thus serve as a sophisticated selection aid. [You should also be aware that in the "real world" a single abstract may incorporate indicative and informative elements, depending on the interests of the intended readers. The type of abstract produced is often determined by its intended readers, the publication content, the journal availability, the language accessibility, and the cost of abstract production. Although for the most part, abstracts are noncritical, abstracts have been known to include a section of critical assessment if the subject warranted one. The length of abstracts depends on the policy of the abstracting service and intended utility. Each service sets specific guidelines for their abstractors. Most abstracts are one-fifth to one-twentieth of the length of the original paper.]
3. A third type of abstract is the critical abstract or review in which the abstractor also functions as an evaluator. For indicative and informative abstracts, the abstractor normally functions as an objective reporter; his or her opinions are carefully excluded. For the critical, or evaluative abstract, the abstractor deliberately injects his opinions and analysis. The value of critical abstracts is highly dependent upon the subject competence of the abstractor, much more so than for the other types of abstracts. Abstracting services do not generally permit critical abstracts because the service cannot be allowed space or time for reply to criticism. Critical abstracts are printed in Applied Mechanics Review and Mathematical Reviews.

INDEXES

An index is a specific kind of tool for finding information. Whether an index is used by a human being or by a machine, its essence is a list of index entries. Each index entry leads to an indexed item somewhere outside the index; for instance, to a record in a database, to a folder in a file drawer, or a book on a library shelf. The entries are in some recognizable order, usually alphabetical. A back-of-the book index is alphabetical by subject and points to a page number(s) in the work with information about that subject:.
HITCHCOCK, ALFRED JOSEPH   14

This entry tells the user that there is information about Alfred Hitchcock on page 14 of the book.

Traditionally, the subject approach to document retrieval is solved by a two-step process. First, each document is analyzed by an indexer according to its subject matter and assigned to one or more concept classes. These concept classes are represented by index terms. In the second step, the indexer chooses the appropriate index terms to represent the concepts identified in the document.

For all practical purposes, subject cataloging is a form of pre-coordinate subject indexing. Subject headings and index terms reflecting the concepts are surrogates for the physical documents. With subject headings, the relationships are established at the time the list is created. However, online catalogs with Boolean search capabilities allow a measure of post-coordinate indexing since the user can specify new relationships by the use of Boolean operators. A file of index terms may thus be arranged, rearranged, manipulated, and searched. As each index term is associated with unique document numbers, documents may be identified. Indexes are not restricted to indexes of subject terms. Index files of authors, titles, report numbers, chemical formulae, and social security numbers may provide other access points if they are useful for retrieval for the users. However, the greatest challenge in document representation is in the creation and maintenance of subject access to documents.

We will look at some of the basic, broad principles behind indexing and then turn to some specifics as to the procedures used in indexing. Subject indexing involves two principal steps:

1. Conceptual analysis -- effective indexing involves deciding what a document is about and how the document is likely to be of interest to a particular group of users. For example, an article may be indexed in several different indexes with different descriptors or terms based on the interests of the targeted group of users. An article on computer imaging of a particular organ of the human body might be idnexed in Index Medicus and in the Engineering Index, but the particular aspects of the article brought out in each index would be different. In the Engineering Index the technical aspects of how the computer technology works would be stressed in the terms selected, while in Index Medicus, the medical aspects would be emphasized.
2. Translation -- involves the conversion of the conceptual analysis of a document into a particular set of index terms usually from some form of controlled vocabulary such as a thesaurus or a list of subject headings.

Now, let's step back and consider the actual design of an index.

There are at least ten fundamental decisions which must be made concerning the design of any index or indexing system. These decisions are:

1. Indexable matter -- within the item or collection of items, what portions should be indexed and what portions should be ignored? Indexable matter refers both to the items within a larger body of materials which are to be indexed and also to the portions of those items which are to be considered in the indexing process. For example, if you were compiling an index to a local newspaper, the newspaper itself is obviously the indexable matter, but would you bothe to index national news? {Probably only as it affects something or someone locally] Would you index advertisements? {Maybe for a brand new business or a going out of business sale to date the birth and death of local enterprises] Very often in the case of periodical indexes, the scope of the index will specifically exclude editorials, letters, advertisements, or reviews of other publications. These are just a sample of the kind of 'indexable matter" decisions which have to be made before jumping into an indexing project.
2. Symbol or concept indexing -- what should constitute the basis of indexing: symbols (e.g., words or pictures in a text) or concepts. It can be said that machines index symbols and humans index concepts. When a human indexes an item, s/he perceives symbols which trigger concepts in the mind. For example, automatic or computerized methods of indexing such and KWIC indexes are based on the actual words in the text, but a human indexer might convert the words used into "higher level" concepts. For example, if a picture showed a small child apparently moving toward a road with a large truck appearing to be headed right where the child's path will intercept the road, a computer can index the actual elements present, but a person will pick up on the concept of potential danger/tragedy about to happen.
3. Depth or exhaustivity of indexing -- how detailed should the indexing be? What should constitute the unit of indexable matter, the subject matter of the work as a whold (as is done in book cataloging) or by the chapter or

article (as in comonly done with periodical indexes) or by the page (as in back-of-the-book indexes)? On the average, how many indexing terms should be used to describe its contents and provide access to it? [3 per page, 3 per article, 3 per book, etc.?]

4. Specificity of indexing -- How specific should the indexing terminology be in relation to the concepts or symbols indexed? I plan to come back to specificity in just a few minutes.

5. Indexing vocabulary -- Should the terms used for index entries be controlled and uniform, or should they be unregulated (left free to reflect the terms used in or by the indexed item or freely chosen by the indexer? In either case, should relations among concepts represented by terms be indicated (e.g., synonomy, genus/species, whole/part, object/operation/agent, or other associations)? Again, we will come back to indexing vocabulary in a few mintues.

6. Surrogation -- How should the indexed item be represented in the index? (how much information is given, what kind of information and in what format?) An index entry must include a description of and reference to the passage, document, thing, person, or organization indicated by the entry. For example, a cataloging record is a surrogate of the entire item being cataloged. An entry in Library Literature is a surrogate for the entire article. The user must have enough information to decide if the item is what is needed and the information necessary to get to the actual item.

7. Record Structure -- How and to what extent should the record representing the indexed items be structured? Structure effects how the files can be searched. Every identifiable element potentially can be searched separately, or in combination with other elements. In order to actually make the specific search, not only must the record structure make each of the elements identifiable, but the searching software must accommodate the search. Search options in an online catalog may be limited not because the MARC records lack structure, but because only a few search options are built into the searching software.

8. Record Display -- How should the record representing the indexed item be displayed in online media? In printed media? (individual record item as opposed to a file display)

9. File Structure -- Direct file or inverted file structure? In machine-readable files, file structure refers to the way the file is arranged within the computer's storage and memory areas. A "direct" file consists of a sequence of item records, one for each indexed item. The sequence may be random or in a meaningful order. The "inverted" file structure involves at least two files, a direct file for the item records and an inverted file for the searchable terms.

10. File Display -- How should files be structured for online access? How should the results of online searches be displayed? How should files be arranged and displayed for print media access? What determines the order on the screen -- alphabetical order (which rules?) or chronological order (earliest records or lateest records displayed first?) or a weighted scheme (attempt to show most relevant items first)?

When a system is evaluated, performance is often expressed in terms of recall (the ability to retrieve useful items) and precision (the ability to avoid useless ones). It is important to select an indexing vocabulary which is more likely to produce the degree of recall and precision desired. Both of these commonly known performance measure have evolved from intensive retrieval testing in the past. The prevailing hypothesis in the 1960s was that indexing language held the key to retrieval performance. Researchers experimented extensively with various indexing languages and several factors in indexing. Several factors in indexing and indexing language were found to exert substantial influences on retrieval. Three important concepts were specificity, exhaustivity, and depth of indexing. Specificity was and is a characteristic of the indexing language. Both exhaustivity and depth of indexing are determined by indexing policy decisions. Each concept is linked with the recall and precision of the indexing language used.

INFLUENCE ON RETRIEVAL:

1. *SPECIFICITY*

In terms of retrieval performance, precision is the percentage of relevant documents contained in the retrieved set. With highly specific index terms, each retrieval set tends to contain highly relevant documents. That is, the precision of the retrieval system increases. Conversely, with a less specific index language, each index term would cover a larger topical domain, not all of which terms are related to the specific area needed. The retrieved set would be larger, and more nonrelevant or marginally relevant documents would be included. Precision of the system suffers. At the same time, as many more documents of less relevance are retrieved, some of these documents may contain information of

pertinence to the topic sought. Specificity of the indexing language is the single most important factor affecting search precision. More generic indexing results in retrieval of greater numbers of items (higher recall).

Obviously, specificity of an index language or thesaurus is determined when the vocabulary is constructed. Once the choice of the vocabulary is determined for the retrieval system, little can be done to change the specificity of the language. Therefore, to begin an indexing project, the choice of an indexing language with the desired level of specificity is an important consideration. It is a challenge to determine exactly how specific the vocabulary needs to be. For most individuals who come to indexing projects after a vocabulary has been in use for some time, there is no easy way to remedy an indexing language with low specificity without a major overhaul and retrospective indexing.

2. *EXHAUSTIVITY*

Indexing exhaustiveness is also a major consideration in terms of recall and precision. Each index term serves as a tag for a theme or concept in the document. If every facet of a paper were indexed and as many as 30 index terms were used to represent the paper, then a search with any one of the index terms would be able to retrieve the paper. Clearly with indexing exhaustivity, a high probability exists that most of the relevant papers as represented by the index term would be retrieved. Exhaustive indexing does insure high recall.

Obviously the treatment of the topic as reflected by the index term in some of the retrieved documents may be less important to the document. Some may even be highly peripheral. If only highly relevant documents on that topic are needed, use of a particular term as a search term would retrieve many papers with only minor mention of the sought topic. High recall from exhaustive indexing often results at the expense of scanning many marginally relevant documents.

Generally, the degree of indexing exhaustivity is proportional to the number of index terms assigned per document. However, depending on the type of publications indexed, it is not a necessary condition. A paper may deal with two or three concepts so that exhaustive indexing could only produce a few index terms for the paper. Suppose a paper represent five different concepts. It is conceivable to index exhaustivity different aspects of three of the five themes with many terms and ignore the other two. In cases of this kind, the number of index terms assigned is not an accurate indicator of the degree of indexing exhaustivity. Therefore, it may be misleading to measure indexing exhaustivity by the average number of terms assigned per document.

3. *DENSITY OF INDEXING*
Often the phrase density of indexing is used interchangeably with indexing exhaustivity. It is a measure of the average number of index terms selected to represent each document. Density of indexing is purely an estimate of exhaustivity. In lieu of indexing exhaustivity, which is hard to do, a more pragmatic measure has been devised. It is believed that although they are not equivalent, experienced indexers can achieve a desired degree of exhaustivity given an upper limit of the number of index terms allowed. Studies have shown that an average of 70 to 80 percent of the total relevant documents in the file can be retrieved if ten terms are assigned for each document. On the other hand, a diminishing return is noted in experiments in which the indexer is asked to assign many more terms. A much greater effort is required to retrieve the last 10 percent of the remaining relevant documents in the file. By assigning an additional 40 to 50 terms per document, 90 percent recall may be achieved. Therefore, a much greater amount of effort must be expended to improve the recall to 90 percent. By requiring many more terms from the indexer, a substantially lowered cost effectiveness is evident.

INDEXING METHODS AND PROCEDURES
The best way to learn how to index a book or a document is to study existing indexes and to use them at length. Indexing is more of an art than a formal, documented procedure such as descriptive cataloging. Although indexes vary widely in their characteristics and quality, a person who examines and uses indexes will gradually learn what an ideal index should be like.

 Good indexing is not a casual clerical job. It is the result of professional activity carried out by people with proper training and experience. We will discuss today some of the procedures and techniques, worked out over the years, that can be learned and followed.

INDEXING PROCESS
 Any indexing involves at least three steps or stages:
FAMILIARIZATION > ANALYSIS > CONVERSION OF CONCEPTS TO INDEX
TERMS

 The first step towards a successful index or set of search keys is familiarization. The indexer must become conversant with the subject content of the document. In order to achieve good consistent indexing the indexer must have a through appreciation of the structure of the subject and the nature of the contribution that the document makes to the advancement of knowledge. The other aspect of the familiarization process is with the particular document itself. To do this a combination of reading and skimming is usually advocated. The parts to be carefully read will be those most likely to tell the most about the contents of the document in the shortest period of time: the title, abstract, summary, and conclusion. Section headings and captions to illustrations or table are also woth attention. If hte particular document happens to have a table of contents, it will also be useful.

The indexer is now ready for the analysis of the document, the second stage prior to index term selection. Depending upon the particular indexing situation, the first step might be to decide whether or not the document is worth indexing based on the material read/skimmed during the familiarization step. If it is judged worthy of being indexed, for any document other than a self-contained book index, the correct bibliographical data according to a consistent format is recorded. Care must be taken to ensure that data is recorded accurately, for the obvious reason that incorrect entries cause the document to become inaccessible. Next a decision as to what parts of the document will be indexed and what parts will be skipped over is made. The human analysis of a document and decisions concerning which subjects are sufficiently significant for indexing is difficult to codify. Some features of the process can be specified, but others rely to a large extent upon experience and intuition. Some topics in documents represent the main theme of the document. Main themes obviously must be represented in indexing, but to what extent need minor or secondary themes be indexed? Some of the subject facets of a document will obviously be basic to the needs of the index users, some will be of marginal interest, and some will be of no importance. Sometimes guidelines are provided that may go some way toward instructing indexers in consistent identification of concepts. Other times it is strictly a judgment call to be made by the indexer, taking into account the needs of the specific organization and patrons. As these decisions are being mad, indexers jot down the concepts, either using words directly out of the text or drawing on their own vocabulary or combination of both.
Where are some good places to look for subject concepts?

1. Title -- assuming the title is indicative of the documents contents.
2. Abstract -- good abstracts are fundamental indicators of subject content. Most of the words in the abstract should heavily convey subject content.
3. Text itself -- Introduction, summary, and conclusions should be consulted. Section headings should be noted along with the first and possibly last sentences of paragraphs since these sentences often carry the message of the paragraph. Note charts, and other illustrative material, methodology, historical and theoretical background. Knowing what to read and what to skim comes from experience.

 The third step involves selection of index terms to match the concepts. (Of course, experienced indexers may merge steps 2 and 3.) This conversion process will differ depending upon the specific type of indexing language used. Some systems use a controlled vocabulary so that an indexer must use a thesaurus to choose index terms. At the other extreme are systems which use free indexing languages which means that any word or term that suits the subject may be assigned as an indexing term.

A variation of free indexing languages is natural language indexing that uses the language of the document. Most natural language indexing is concerned with machine assignment of terms, and is based upon the language of titles and abstracts. Selection is simple and there is no need for scanning and analysis of documents. There is still active debate as to whether natural language indexing leads to effective retrieval.

LIMITATIONS OF TERM INDEXING

Unfortunately, despite the dominance of human indexing (as opposed to automatic or machine indexing), very little is known about the intellectual task of indexing. Over the years, there have been numerous studies on various aspects of indexing. At one time, the quality of indexing was thought to be the dominant factor influencing the retrieval performance of information retrieval systems. Despite intensive investigation of how one indexes, very little insight has been gained as to how one transforms text content into index entries. In fact, indexing manuals often avoid the issue entirely.

Although the understanding of the process of indexing is lacking, serious defects in this approach are well known. Problems such as inter-indexer consistency have been documented, i.e. different indexers tend not to index with the same index terms. Often to aid in retrieval, index terms in online systems are also supplemented by keywords found in titles and abstracts. In recent years, many full-text retrieval systems have been developed in which little or no human indexing is needed. Each document is automatically indexed by every nontrivial word in the text or title or abstract. A few systems even allow weighting of index terms to further aid retrieval.

Professional Societies: American Society of Indexers

For those of you considering indexing as a major part of your professional career, there is a professional society called the American Society of Indexers, that you may be interested in joining in the future. The society was founded in 1968 and at that time affiliated with the [British] Society of Indexers, and, thus, shared its journal, The Indexer until 1998.

Ten years after the founding of ASI, the H. W. Wilson Co. established an annual award for excellence in indexing, which is administered by ASI. A monetary award is given to the compiler of the best index to a monograph, and a certificate goes to the publisher. The criteria for this award constitute an important standard for the evaluation of book indexes.

The international journal The Indexer is not a journal of indexing research -- articles of that nature are more likely to appear in the Journal of the American Society for Information Science (JASIS). Rather, it contains state-of-the-art reviews on computer-assisted indexing, descriptions of indexing projects, think pieces, as well as a humorous column that excerpts comments on indexes from book reviews.

One of the main purposes of the Society is to convey the importance of quality indexing. To this end, ASI produces brochures with index evaluation checklists, exhibits at publishing conferences and distributes complimentary copies of its annual Register of Indexers.

Indexing has recently been profiled in several publications as a "work-at-home" career which has generated publicity for ASI. ASI members are predominantly self-employed and create indexes mainly to monographic publications.

Other professional societies to which information professionals belong who are interested in indexing and abstracting are: ASIS, particularly to the Classification Research SIG; ALA Association for Library Collections and Technical Services (ALCTS) [differences between cataloging and indexing essentially being one of degree of analysis]; National Association of Abstracting and Information Services (NFAIS) [an organization of primarily corporate members who are database producers]; and the Special Libraries Association [indexing rather than full blown cataloging is often utilized by special libraries].

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	234	document and replac$3 and (document adj (extract or abstract or summary))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2005/08/01 14:09
L2	9	document same replac$3 same (document adj (extract or abstract or summary))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2005/08/01 14:17
L3	152	((creat$4 or compil$6 or gather$3 or assembl$4) with (database$1 or data adj (store$1 or base$1 or repositor$3 or warehouse$1))) and (document adj (extract or abstract or summary))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2005/08/01 14:19
L4	2	((creat$4 or compil$6 or gather$3 or assembl$4) with (database$1 or data adj (store$1 or base$1 or repositor$3 or warehouse$1))) with (document adj (extract or abstract or summary))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2005/08/01 14:24
L5	14	((sav$3 or stor$3) with (database$1 or data adj (store$1 or base$1 or repositor$3 or warehouse$1))) with (document adj (extract or abstract or summary))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2005/08/01 14:27
L6	95	(sav$3 or stor$3) with (document adj (extract or abstract or summary))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2005/08/01 14:30
L7	13	((sav$3 or stor$3) and (conver$4 or creat$3)) with (document adj (extract or abstract or summary))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2005/08/01 14:36
L8	13	((conver$4 or creat$3) with (document adj (extract or abstract or summary))) and ((sav$3 or stor$3) with (document adj (extract or abstract or summary)))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2005/08/01 14:56
L9	1	document adj abstracting	EPO; DERWENT; IBM_TDB	OR	OFF	2005/08/01 14:57